

THE ENGLISH WORD : A Critical Survey of Some Aspects
of Lexicography and Lexicology in the English Language

Thomas Burns McArthur

Doctor of Philosophy
University of Edinburgh
June 1976



Preface and Acknowledgements

This study attempts two things: to provide a wide-ranging review of English-language lexicography, and to inquire into the nature of English words. It stems from a belief that these two are complementary aspects of one inquiry.

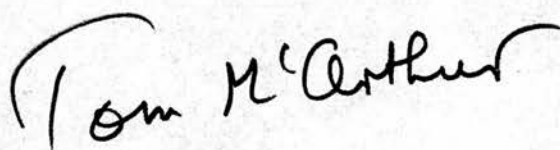
It should be noted that the bibliography stops at publications in 1973. Because of the scope of the inquiry — ranging over lexicography, linguistics, anthropology and related disciplines — it was thought that this cut-off point was realistic in order to allow the work to finish.

I would like to thank David Abercrombie, Professor of Phonetics at the University of Edinburgh, without whose constant interest and sympathetic help the work would not have been brought to a conclusion. Advice on Chapters 1 and 2 from Professor Randolph Quirk of University College, London, is also gratefully acknowledged. I should like also to thank the Co-ordinating Indexing Group of the Association of Librarians (ASLIB) for the opportunity of addressing three conferences (1973, 75 and 76) on aspects of the work in relation to information science; Mary Abercrombie for her advice in Chapter 3, in relation to the pioneering work of Isaac Pitman; Dr Robert Boyer for his help in the computer programming of material described in Chapter 5 and the appendix; the lexicographic staff of Longman, Harlow, especially Paul Procter, for the stimulus and practical opportunities of developing ideas outlined here; and Gordon

Jarvie of Collins, for his practical help in preparing material relating to Chapter 5 and the appendix for publication as the Patterns of English series Books 1 to 4. Acknowledgement should also be made of Dr Patrick Allen and Mr Paul van Buren of the Department of Linguistics, the University of Edinburgh, for supervisory work on the study in its first two years.

I would also like to acknowledge with gratitude the determined support of my wife Peri, especially at those times when it seemed unlikely that the work would be completed; and also the excellent typing of the chapters and appendix by Mrs Pat Drummond, who has made the whole thing look worthwhile.

Many have helped. The support and advice is theirs; the faults are mine. The thesis is, in all respects, my own work.

A handwritten signature in black ink that reads "Tom McArthur". The signature is written in a cursive, slightly slanted style. The first letter "T" is large and loops around the "om". The "Mc" is written in a simple, connected way, and "Arthur" follows with a long, sweeping tail that extends to the right.

Tom McArthur

Longniddry, June 1976

Abstract

Chapters 1 and 2 of this study trace the development of English-language lexicography through two distinct traditions: the glossary-to-dictionary line and the less well-documented vocabulary-to-thesaurus line. It shows that compilers have never explicitly formulated a theory of 'the word'. Instead, they developed many practical techniques for listing, retrieving, defining and illustrating language items which (they assumed) every educated person automatically knew were words.

English-language lexicography began as an exercise in translation between Latin and English, but, as Latin material was absorbed into English lexis (from the Renaissance onwards), lexicography became an apparently unilingual activity. The evidence indicates, however, that despite appearances the original bilingualism remains with us, translation being now intra- instead of inter-linguistic.

Chapter 3 reviews the counters of words from the early 19th century on, surveying the objective counters (such as Thorndike), and those who sorted words subjectively (such as Palmer), along with a parallel logico-semantic approach (Ogden's Basic English). Here again there was a lack of explicit theory, and the counts were less than successful because of a failure to be clear on what to count and where to go for the basic data. The word-counters were also

inevitably drawn by the logic of their work into the business of compiling dictionaries.

In Chapter 4 linguistics is seen as having made few explicit attempts to define words, nevertheless frequently appealing to our inherent assumptions about what words are, so that such units as 'morphemes' could be postulated and put to work.

Chapter 5 synthesizes elements in the earlier chapters. A typology is offered for the English word, and a distinction established between words and lexical bases for word-formation. A number of devices are proposed as useful in any theory of words, including structure formulas, a root-and-base distinction, holism, derivational paradigms and compounding patterns. The theoretical position adopted derives from the work of many linguists, but in particular from Vendryes, Sapir, Ross, Entwhistle and Marchand. A review is also made of the problems relating to semantic analysis, as undertaken by the American cognitive anthropologists and by Lyons. It is proposed that English lexis, historically and functionally, is polysystemic, a composite of two (Vernacular and Neo-Latin) 'streams'. These interact in the living language to provide parallel morphologies and reservoir areas for word-coining.

Chapter 6 is an attempt to demonstrate how the principles of Chapter 5 can be applied to a specific area: suffixal word-formation in English. It adopts the polysystemic approach, is both diachronic

and synchronic, and uses derivational paradigms, paraphrases and glosses to create a new model of productive suffixation. It is argued that the lexis and morphology of English cannot be adequately understood without recourse to polystemic models of this kind.

Preface and Acknowledgements

Abstract

1	The Glossary Tradition in English Lexicography	1
1.1	Introduction	1
1.2	Gloss, Glossary and Vocabulary	1
1.3	Latin and English : Distinct Languages	4
1.4	Latin in English : The Distinction Blurred	7
1.5	The Legislative Urge	14
1.6	Johnson : Consistencies and Contradictions	20
1.7	Orthography and Orthoepey	24
1.8	The Encyclopaedia	28
1.9	Webster : Consistencies and Contradictions	31
1.10	The Dictionary on Historical Principles	35
1.11	Murray's Dictionary	38
1.12	Conclusion	42
2	The Vocabulary Tradition in English Lexicography	44
2.1	Introduction	44
2.2	Models of Cosmic Order	45
2.3	Discriminating Synonyms	55
2.4	Roget's Thesaurus	62
2.5	Words, Ideas and Thoughts	69
2.6	The 'Sacred' Thesaurus	73
2.7	March's Thesaurus - Dictionary	75
2.8	Conclusion	81

3	The Vocabulary-Control Movement in the English Language	84
3.1	Introduction	84
3.2	The Special Word List	85
3.3	Thorndike : the Objective Approach	88
3.4	Palmer and West : the Subjective Approach	95
3.5	Ogden : the Logical Approach	100
3.6	The Carnegie Report and the General Service List	106
3.7	Commentators on the Lists	111
3.8	Conclusion	117
4	Linguistics, the Morpheme and the Word	120
4.1	Introduction	120
4.2	The Three Ages of Etymology	120
4.3	Skeat : the Root of the Matter	125
4.4	Historical Processes and Grammatical Processes	128
4.5	Lexical Word and Grammatical Word	131
4.6	The Problem of the Morpheme	134
4.7	American Structuralism	139
4.8	Chomsky and the Syntactic Machine	144
4.9	Halliday : Word and Lexical Item	150
4.10	Conclusion	

5	Towards a Theory of the English Word	
5.1	Introduction	156
5.2	A Typology of Words	157
5.3	Morphology and Syntax	165
5.4	An Indo-European Word Formula	172
5.5	The Phenomenon of Holism	175
5.6	Derivational Paradigms	183
5.7	The Compounding of Words	189
5.8	Language, Lexis and Conceptual Universes	195
5.9	Structural Semantics	204
5.10	Parallelism in English Lexis	209
5.11	Models of English Lexis	221
5.12	Conclusion	232

6	Some Derivational Paradigms for Suffixation in English	237
	Introduction	237
	Paradigm 1 Basic Deverbal Pattern	242
	Paradigm 2 Basic Deadjectival Pattern	250
	Paradigm 3 Basic Animate Denominal Pattern	254
	Paradigm 4 Basic Inanimate Denominal Pattern	257
	Paradigm 5 Denominal in <u>+al</u>	259
	Paradigm 6 Denominal in <u>+an</u>	264
	Paradigm 7 Deverbal in <u>+ion/ive</u>	267
	Paradigm 8 Restricted Denominals	271
	Paradigm 9 Denominal in <u>+ic</u>	274
	Paradigm 10 Nominalizing in <u>+ic(s)</u>	281
	Paradigm 11 Denominal in <u>+oid</u>	284
	Paradigm 12 Denominal and Deadjectival in <u>+ify</u>	286
	Paradigm 13 Denominal and Deadjectival in <u>+ize</u>	291
	Paradigm 14 Denominal in <u>+ism</u> , Uncountable	295
	Paradigm 15 Denominal in <u>+ism</u> , Countable	299
	Bibliography	301

1 The Glossary Tradition in English Lexicography

1.1 Introduction

Dictionaries have a sanctity similar to Holy Writ. Samuel Johnson may have characterized the lexicographer as a 'harmless drudge', but this drudgery puts its victim in the company of the prophets. People do not distinguish readily between this or that dictionary, or between better or worse dictionaries, but conflate them (with that prestigious definite article) in company with The Bible, The Koran and The Bhagavad Gita.

Weekley (1924 (1962:9)) saw something touching in the belief that 'the dictionary, like photography, does not lie', and observes that 'almost the only individual to approach the sacred book in the spirit of a doubter is the lexicographer himself.'

The Dictionary, that spurious monolith, has a long and convoluted history. Its progress within the English language has been mapped by such historians of lexicography as Murray (1900), Mathews (1933), Starnes and Noyes (1946) and Whitehall (1958), and in this chapter I shall try to review and supplement these commentators, to establish the stages in the growth of the lexicographic tradition in English and to consider how the makers of dictionaries viewed the English word.

1.2 Gloss, Glossary and Vocabulary

General agreement assigns the origins of English-language lexicography to the 7th century AD, to the time of St Isidore of Seville. The

Latin language was the vehicle of culture and religious authority at that time, and young aspirants to clerical careers in Western Europe had to master it, whether they were Provençals, Burgundians, Celts or Anglo-Saxons.

Manuscript books in the 7th and subsequent centuries were scarce and valuable, but students and teachers did not scruple to write on them. In all literate cultures, users of books have tended to insert the 'meanings' of difficult words in the margins and between the lines, and it is in such practices that the great lexical works of English had their start. A hard Latin word would be 'explained' by means of a simpler one or by a vernacular term, usually put in just above the word in question. The Greek word for a difficult or special term was glossa, and in Neo-Latin the meaning of this word was extended to include the explanation of a difficult term as well as the term itself. In consequence, these mediaeval notes are referred to today as 'inter-linear glosses'.

By the 8th century, glossae collectae were being prepared by copying the hard words and their explanations into short lists. Such lists, put together into larger lists, were called glossaria and with their recopying, development and increasing length some kind of ordering for purposes of retrieval became necessary. Crude alphabetic order was introduced and later lists were re-cast with increasing sophistication. Detailed alphabetic ordering came slowly. Thus, the Leiden Glossary is the most primitive: a collection of lists each with the name of the treatise from which items had been extracted, the items left in order of extraction. The Epinal Glossary moves to first-letter order, with, for example, 550 entries under 'A'. The Corpus

Glossary (Cambridge, compiled about 725 AD) contains material from these and others, with second-letter ordering, which allows, for example, 95 unordered words under 'Ab'. There is a 10th-century glossary in the British Museum (Harl. 3376) with third-letter order, but no contemporary evidence of further refinement.

The alphabetization which a 20th-century user might take for granted was therefore a slow acquisition, as was columnar listing. Up to the 16th century, it was optional whether entries would be listed horizontally (the explanation above the explained term, as in the inter-linear glosses) or vertically (the explanation to the right of the explained term).

Most commentators follow Murray (1900:9f) in assuming that these Latin glossaries were a major source of lexicography, but not the only source. The other was pedagogical, from the classroom rather than the reader. It was necessary to learn the words (vocabula) as well as the grammar of Latin, and although this was largely an oral undertaking, lists were compiled of parts of the body, animals, heavenly objects, geographical features etc. Such a list was known as a vocabulary and was distinct in style and use from a glossarium. Alphabetization was of little interest in such groupings of words under semantic heads.

Glossary and vocabulary were employed for complementary work and were often interwoven; but as becomes clear in Starnes and Noyes (SN: 197ff) the two represented distinct traditions which might converge, but which tended to keep separate courses. This separateness is important, and should not be pre-judged by labelling the two traditions

by their current end-products. Better instead to refer to them as the glossary tradition on the one hand (characterized by explanation and alphabetization) and the vocabulary tradition on the other (characterized by grouping into fields of interest). The history of lexicography tends to emphasize the first at the expense of the second, but the second tradition is also important. We shall return to it in the second chapter.

1.3 Latin and English: Distinct Languages

The ancient glossaries not only became fuller and more orderly as time passed, but also began to emphasize the vernacular rather than the Latin. In the Epinal Glossary the English words are relatively few throughout, in the Corpus Glossary they are more numerous, and in 10th-11th century lists (glossary or vocabulary) we find true Latin-to-English translation. The bilingual principle had emerged, as opposed to a rule-of-thumb approach offering anything Latin or English which might help the student to understand.

Both Murray and Mathews maintain that the Norman Conquest halted the development of lexicography, creating an unstable linguistic situation for over three centuries. Certainly, the compilers Wright-Wülcker (1857 and 1884) show only two slender vocabularies between 1066 and 1400 - but after that year output becomes prodigious, with English more and more in the ascendant. The development of both lexical traditions was assured not only by educational development, but by the invention of the printing press.

The Medulla Grammatices is an anonymous Latin-to-English glossary of the mid-15th century. It was never printed but served as the basis

for the later Ortus (= hortus, 'garden') Vocabulorum of 1500, the first printed Latin-to-English wordbook.

A Dominican friar called Galfridus Grammaticus (Geoffrey the Grammarian) appears to have been the first to make the revolutionary switch-round in language order. About 1440, he produced an English-to-Latin glossary, called the Promptorium Parvulorum (sive Clericorum), or 'storehouse for little ones (and clerics).' Some 10,000 nouns and verbs were provided in two lists, with other adjectives and particles thrown in where they seemed appropriate. In 1499 Pynson issued the first printed version, and with the Ortus a commercially successful tradition of bilingual glossaries was set underway.

The compilers of the various lists up to the 16th century had no specific word for what they were doing. They sought elegant metaphors to express their aims. Wynkyn de Worde, the printer of the Ortus, justified its title by saying that it was like a garden filled with wholesome flowers, fruit and herbs. The Promptorium (= promptuarium) suggests a store from which words can readily be plucked for use. As will be seen, other terms abounded: thesaurus (= treasury), manipulus (= handful), dictionarium (= phrasebook; collection of dictions, sayings), alvearie (= beehive), bibliotheca and library (= place of books, knowledge etc) and the vivid abecedarium, saying so much for the system of ordering.

Thomas Elyot chose 'dictionary' as the name for his wordbook, a compilation in 1538 which had notable success. Thomas Cooper of Magdalen College, Oxford, brought out augmented versions of Elyot (1548, 1552 and 1559) and using additional continental material absorbed

the Elyot list into his own larger Thesaurus Linguae Romanae et Britannicae (1565). Cooper was in turn abridged by Thomas Thomas in his Dictionarium Linguae Latinae et Anglicanae (1588) which was successful well into the 17th century and was to have immense repercussions on English lexicography. The Elyot-Cooper-Thomas line is one of the strongly influential traditions within the glossary tradition at large, giving us not only techniques of definition and coverage, but also the important names 'dictionary' and 'thesaurus', although not yet used in anything like their current senses.

The impetus of trade, travel and exploration in 16th-century Europe promoted not only the Latin language but also the vernaculars which developed in its shade. In 1514 a Londoner, John Palsgrave, was commissioned to teach French to Henry VIII's sister Mary. The treatise which he produced to help her was published in 1530 as Esclaircissement de la langue francoyse, combining features of the grammar, the teaching manual and the vocabulary. A profusion of bilingual works followed, and English could be studied by looking at how its words were used in comparison with French, Spanish, Italian, Welsh and Latin words.

While this interest in other languages flourished, a controversy was gathering regarding the nature and future of English itself. It is difficult to date its beginning, but Mathews (14) notes that 'mutterings of protest were beginning to be heard' by 1491, the year in which Caxton the printer died. He had characterized his people and their language as under the domination of the ever-changing moon, and the dispute concerned how much change and transfusion from abroad English could take without being corrupted beyond redemption.

1.4 Latin in English: the Distinction Blurred

The controversy was between the innovators and the purists, those who willingly imported foreign words, and those who resisted the impertinence. The Renaissance was characterized by the influence of Neo-Latin on the emerging vernaculars of western Europe; the vernaculars were subject to an immense cultural pressure from Neo-Latin, the language of learning and international debate. Significantly, the purists of English objected more to the Neo-Latin intrusion than to the lesser inrush of Hebrew and non-European words. Not only did they resist Neo-Latin, but they also objected to new coinages within Neo-Latin. There were those who argued that Neo-Latin was itself a corruption of Latin's Ciceronian glories.

The controversy lasted some 200 years, and in the opinion of Victor Grove (1950) in The Language Bar, the clash of Neo-Latin and original English has never been satisfactorily resolved. Many in his opinion still feel deprived and shut-out through an inability to cope with 'high-brow' Latinate vocabulary; the Elizabethan Wilson, in his Art of Rhetorique (1553) may still speak for many when he says: 'Some seeke so farre for outlandishe English, that thei forget altogether their mothers tongue.'

Many Renaissance figures, while venerating Neo-Latin, encouraged the growth of their own vernaculars, the result being an infusion of the vocabulary of the cultural language into the emerging language both in England and Scotland. Although his dictionary shows him a confirmed Latinist, Sir Thomas Elyot wrote in English very successfully. He considered his own language in need of supplementation where it fell

short, and borrowed and coined freely to this end (SN:6). At the same time, however, he felt he was being moderate: he could congratulate himself on Henry VIII's opinion that The Governour was not 'therby made derke or harde to be vnderstande' (Mathews: 15). Excessive Latinization of English was however severely mocked, as we see from Shakespeare's *Holofernes*.

Neo-Latin as the dominant cultural language of Christendom can be compared to Classical Arabic in the Middle-East and to Sanskrit in the Indian subcontinent. The local vernaculars (whether related historically to the dominant medium or not) have been unable to resist lexical intrusion from Arabic and Sanskrit and Latin, so that strata of specialized vocabularies form within them. Similarly, modern Indian vernaculars like Hindi and Marathi are subject to extensive intrusion from English, and feelings in India today about innovation and purism are similar to those of the Elizabethans.

By the end of the 16th century, Neo-Latin was both outside and inside English, and in Whitehall's phrase (158), 'a race of *Holofernes* pedants' preferred the Latinate to the English, using 'Latino-Greek polysyllabics in a Latino-English syntax.' Sometimes emphasis on the more outlandish constructions obscures an important point: that the intrusion was on many stylistic levels, not just in academic pedantry. George Pettie (1586) observed of the learned words: 'If they should be all counted inkpot tearmes, I know not how we should speak anie thing without blacking our mouths with inke.'

Whatever the merits of abetting or resisting the influence of the dominant medium, schoolmasters had the problem of coming to terms with

the swirl of words. They were interested in some kind of standardization of spelling, some yardstick for grammar, some kind of clarity about vocabulary. And they began to suggest a dictionary of some kind. A grammar in 1594 by one 'P.Gr' contained a dictionariolum, a vocabulary of English words with their Latin equivalents, while Richard Mulcaster had in 1582 written a marginal note: 'a perfit English dictionarie wished for.' He even compiled some 8,000 words worth defining in such a dictionary. Edmund Coote (The English Schoole Master, 1596) created a model with a collection of some 1,500 'hard words' and their simpler definitions, and provided the foundation for a small but momentous octavo volume in 1604: Robert Cawdrey's A Table Alphabeticall (of Hard Words), which discussed borrowings from Latin, Greek and other tongues, 'gathered for the benefit & helpe of Ladies, Gentlewomen, or any other unskilfull persons.'

Cawdrey's 3,000-item book was followed by John Bullokar's The English Expositor (1616), Henry Cockeram's The English Dictionarie (1623), Thomas Blount's Glossographia (1656) and Edward Phillips's The New World of English Words (1658). These are the tradition of the 'hard-word dictionaries', their popularity wondered at by later generations of critics. Like the bilingual glossaries, they were a commercial success, through to the 13th edition of Bullokar in Dublin in 1726, a run of over a century. The social history of the compilers and their books is fascinating (of SN, 13ff), but our immediate concern is their incipient lexicographic method. They can be said to have set in motion two techniques which are still powerful and indeed expected in dictionaries:

- 1 fixing the spelling of a word
- 2 giving a synonym for it.

The following 15 hard words and definitions taken casually from Cawdrey's Table Alphabeticall show a spelling close to what is still acceptable, along with an interesting contrast of Latin and Anglo-Saxon words (the Anglo-Saxon material in the explanation column being underlined here for convenient display):

abrogate	<u>take away</u> , <u>disanull</u> , <u>disallow</u>
acquisition	<u>getting</u> , <u>purchasing</u>
aggravate	make more grievous, and more <u>heavie</u>
direct	guide or rule: <u>right</u> , <u>straight</u> , also to order
disadvantageous	<u>hindering</u> much
gratulate	to be <u>glad</u> for anothers <u>sake</u>
impose	<u>lay upon</u> or <u>put on</u>
impotent	<u>weake</u> , feeble
edifice	<u>building</u>
magnitude	<u>greatness</u>
manuring	<u>dung</u> , <u>tilling</u>
paucitie	<u>fewness</u> or <u>smale</u> number
prompt	<u>ready</u> , <u>quicke</u>
ruinous	<u>ready</u> to <u>fall</u>
ruminate	to <u>chew over againe</u> , to studie <u>earnestlie</u> uppon

This contrast is much the same as the kind of contrast offered in the Latin-to-English dictionaries of a few decades earlier. It is a kind of translation, but intra- rather than inter-linguistic. We can return to this point in a moment: the significance of Cawdrey can be highlighted another way. If we consult the Random House Dictionary (RHD, 1966) as the most recent unabridged dictionary of English, and take each of these hard words of Cawdrey in turn, the similarity of definition in every case (even to the item gratulate) is quite remarkable.

The definitions are longer, better organized, supported by etymologies and subdivided into senses, but they are based upon the same kind of Latin-versus-Anglo-Saxon translation. Suffice it here to cite four only (the Anglo-Saxon material again being underlined):

- edifice a building, esp. one of large size or imposing appearance
- impose 1 to lay on or set as something to be borne, endured, obeyed, fulfilled, paid etc.
 2 to put or set by or as by authority
- prompt ready in action; quick to act as occasion demands
- ruminate vi 1 to chew the cud, as a ruminant 2 to meditate or muse; ponder vt 3 to chew again or over and over
 4 to meditate on; ponder

Here we have not just an indication of consistency in definition within the English language (one assumes that the editors of the RHD were not referring to Cawdrey for their definitions!) but also evidence of an on-going tradition of internal translation, more usually described as 'definition by synonym'.

Starnes and Noyes (21f) implicitly establish this process of external-becoming-internal-translation where they examine Cawdrey's successor Bullokar, whose adoption and adaptation of Latin words was wholesale. They correlate material in the Thomas Latin-to-English glossary and Bullokar's Expositor, thus:

Thomas, 1606 edition

Alacritas... Cheerfulness, liveliness, courage, readiness

Catalogus... a rehearsall in wordes, or table in writing of the number of things, a roll, a bill, a scroll, a catalogue: also a register of proper names.

Rumino... To chew the cudde as neate doe; also to call to remembrance and consider with ones selfe, to study and thinke upon matters.

Bullokar's Expositor, 1616

Alacritie. Cheerfulness; courage, quicknesse.

Catalogue. A roll, a bill, a register of names or other things.

Ruminate. To chew over againe as beasts doe, that chew the cud: wherefore it is often taken to studie and thinke much of a matter.

Cockeram (whose book is the first 'dictionary' of English so named) took the matter further. The volume has three parts:

- 1 an alphabetic list of 'hard words' explained by 'vulgar' words
- 2 an alphabetic list of 'vulgar' words with corresponding 'hard words'
- 3 a glossary of the names of gods, monsters, birds, people etc.

The first two are a bilingual dictionary inside English, to help the user convert to and from planes of style: the elegant/refined and the vulgar. Again, the evidence of his intention and sources is plentiful (cf SN: 31ff). His first section relies heavily on Thomas, as Bullokar did (natatio becoming natation, 'swimming'; necessitudo becoming necessitude, 'kime, alliance in bloud' etc). Additionally, both first and second sections leaned even more heavily on the English-to-Latin compilation of Rider (1589) and Rider-Holyoke (1617), thus:

<u>'vulgar' English</u>	<u>Rider-Holyoke Latin</u>	<u>Cockeram's hard word</u>
to burden	onero	onerate
burdenous	onerousus	onerous
to charme, enchaunt	incanto	incantate
a charme, enchantment	incantatio	incantation
cheapnes	vilitas	vility
to chide or rebuke	iurgo, obiurgo	objurgate
childishness	pueritia, puerilitas	puerility

Starnes and Noyes comment: 'It is obvious that Cockeram found most, if not all, of his so-called "vulgar" words in the English-Latin section of the Rider-Holyoke dictionary; and the "more refined and elegant" terms represent Cockeram's attempts to Anglicize Rider's Latin equivalents of the English. This process is significant in thus introducing a great many Latin words into the English vocabulary. Some of these were already current; others were not' (33).

The compilers of the hard-word dictionaries were not therefore recorders of usage, as lexicographers are nowadays assumed to be. They were manufacturers, transferrers of words wholesale, animated by a belief in the value of what they were doing. The later charges of plagiarism and corruption would have made no sense to them, and it is small wonder that purists had little success in stemming the Neo-Latin tide, when lexicographers as well as schoolmasters and writers were ranged against them. It seems reasonable to suppose that these compilers (themselves well-versed in Neo-Latin) did not draw a sharp line between Neo-Latin vocabulary and the refined, elegant or academic vocabulary of English. It was the same vocabulary, convertible from international currency to local use by means of a few well-tried morpho-graphological rules. At that time the floodgates were at their widest, and the entire lexis of Neo-Latin was available for adaptation to use within an English grammatical framework.

It is not possible to measure just how consciously English was Latinized at that time, but here we have clear evidence of conscious systematic transfer within the general flux. Cawdrey and his successors knew exactly what they were doing, and were a commercial success in doing it.

They possessed one consistent trait: they sought to broaden the base of the educated Elect. Their books were for the non-scholarly, the artisans, the merchants, the wives of gentry and bourgeoisie, and all other aspirants to education and elegance. They offered access to High English for those who felt they only knew the Low variety. Bulloker is quite explicit about it in his preface to the 'courteous reader':

'And herein I hope such learned (professors) will deeme no wrong offered to themselves or dishonour to Learning, in that I open the signification of such words, to the capacitie of the ignorant, whereby they may conceive and use them as well as those which have bestowed long study in the languages, for considering it is familiar among best writers to usurpe strange words... I suppose withall their desire is that they should be understood.'

1.5 The Legislative Urge

The hard-word dictionary of the 17th century was the direct and logical descendant of the interlinear gloss. In it the alphabetic principle is firmly established, although the idea of one composite list was not insisted upon. A dictionary might contain one, three or more separate alphabetic lists, under subject headings, a practice still adopted by some publications (as with the Readers Digest Great Encyclopaedic Dictionary, Vol. III, 1972). Attention was devoted to difficult and strange words flowing in under the pressures of the Renaissance, but even with such an emphasis the compilers were not unaware of spelling problems, indicators of pronunciation, grammatical notes, etymology, cant expressions, dialect words, proverbial sayings and such other scraps which would one day have lexicographic traditions of their own.

The work of these men is often reviewed in a spirit of amusement, commentators enjoying so much to quote their quaint and more risible entries that the bulk of acceptable material is obscured. Murray, Mathews and others have tended to divide the Neo-Latin innovations of these compilers into two categories:

- 1 the astonishing-amusing-unsuccessful
- 2 words now accepted in the vocabulary.

This is a dichotomy dictated by hindsight, not one which had any meaning in their day. Murray notes (18) that in Elyot's time no one could have predicted that the word 'dictionary' would win the day as the term for wordbooks; in the same way, we can say that none of the hard-word compilers could have foreseen that some words would fail and others succeed.

Commentators suggest in quoting items like bubulcitate ('to cry like a cow-boy') that oddity and length were enough to make them extinct, like the Brontosaurus. This is a kind of Darwinian view, and of course the words are funny to us: but other monsters have survived despite everything, such terms as bibliography, incontrovertibility and hallucinatory which seem to have much in common with the extinct terms. Conversely, many small and tidy words have failed in the race: vility, nexible, and naustible did not survive while algid and ablepsy have. One man's monstrous term may be another man's everyday companion, and one has to take care in pronouncing a word safely 'extinct'. In some remote discipline it may still have work to do.

Internal translation (the form which definition by synonym commonly takes) is an established practice in English, not only among dictionary-makers but also in the general educated world, and especially in explaining harder words to younger people. Latinate verbs like abrogate and ruminate were explained in Cawdrey's Table by means of the phrasal Anglo-Saxon verbs 'take away' and 'chew over'. Dictionaries still go in for circularities like feminine explained as 'womanly' and womanly explained as 'feminine'. In a two-tier vocabulary like English, this kind of thing is the easy way out of a problem. It becomes rather daunting however when bull is explained by the gloss male bovine quadruped.

an explanation perhaps with some scientific merit, but serving generally to 'make derke or harde to be vnderstande'.

The dispute between purist and innovator gave way to a different emphasis, one which began to emerge in the mid-1600s. The Italian and French Academies had impressed English writers by the verve with which they had set out to 'fix' their respective languages. In 1664, the Royal Society formed a committee 'for improving the English tongue' and initiated an interventionist approach to the English language and its lexicography. When, after 20 years of preparation and 40 years of work, the French Academy produced its non-alphabetic Dictionnaire de l'Academie Francaise, envy was felt across the Channel.

The French achievement was a serious challenge to Dryden, Addison, Swift, Pope and the other proponents of a perfectible English. Various, they deplored corruptions in the tongue while setting up in their own writings literary models which others might emulate. That period seems to have combined a Stoic fear of corruption with a sense that men were participating in a new Golden Age, but one which might pass away too quickly if something were not done to halt decay. When the French dictionary was re-cast in 1718 in alphabetic form, discussion in the coffee houses and salons turned to an English equivalent. Addison appears to have planned a dictionary of quotations; Ambrose Phillips outlined a plan for a dictionary including orthography, etymology, definitions both literal and figurative, idioms etc; Pope approved the idea and probably drew up a list of quoteworthy authors, but no society emerged to do what the Forty Immortals had done for French.

In the meantime the hard-word dictionaries carried on, the work of schoolmasters and amateurs who called themselves 'bibliophiles'.

And just as Cawdrey's little volume turned one kind of lexicography into another, so a small book in 1702 changed the emphasis from hard words to the general mass of English words. This was 'J.K.'s' New English Dictionary, which in the new spirit declared itself in favour of 'genuine' English words as 'used by Persons of clear Judgment and good Style'.

This J.K. was probably the John Kersey who later made his name both as a reviser of Phillips (The New World of English Words, 1658, revised 1706) and with his own Dictionary Anglo-Britannicum in 1708. He has a number of 'firsts' to his credit:

- 1 blending the hard words and the general vocabulary in one list
- 2 bringing out the first folio 'universal' dictionary and then an abridgement of it
- 3 giving special attention to the definition of technical terms.

His various works, re-issued at regular intervals, were not supplanted by anything new for thirteen years.

Kersey's successor was Nathan(iel) Bailey, a schoolmaster and prolific writer of various educational works. Among his dictionaries two stand pre-eminent: his Universal Etymological English Dictionary (1721) and the folio Dictionary Britannicum (1730). The first is the dominant wordbook of the 18th century, running to 30 editions up to 1802, while the second provided Johnson with a foundation for his work.

Whitehall (159) asserts Bailey's place in lexicography as follows: 'The position of dictionary pioneer, commonly granted to Johnson or to Noah Webster, belongs in reality to one of the few geniuses lexicography ever produced: Nathaniel Bailey.' Starnes and

Noyes (98) are not so sure: 'The spotlight of fame which has long been focused on Johnson and has recently spread to his immediate predecessor, Bailey, has unfortunately thrown Kersey into the shadow. Kersey was, however, a notable pioneer, rejecting outmoded material and methods, working towards modern concepts, and in general playing his role of lexicographer with responsibility and intelligence.'

Genius and intelligence are difficult to assess, and it might be as well to spread the spotlight far enough to include all three, to propose a Kersey-Bailey-Johnson tradition similar to the bilingual Tudor tradition of Elyot-Cooper-Thomas. Evidence abounds to suggest that all three men helped to create the 'modern' dictionary.

The idea of a comprehensive English dictionary grows through Kersey and Bailey, parallel to the cry in the coffee houses for something to match the French. In Bailey certain clear fundamentals appear:

- 1 a necessary place for etymology, but bracketed so that the user may by-pass history if he wishes
- 2 elementary help with syllabification, accent position, spelling and pronunciation
- 3 woodcut and copperplate illustrations
- 4 appended proverbs, idioms, current sayings and illustrative quotations.

Bailey was not an academic: he was a commercial compiler, interested in entertaining as well as informing his readers. This can be seen in the entry for the verb live:

To live (libhan or leofan, Sax. lefwe, Dan. lefwa, Su. Leben, Du. and L. G. leben, H.G.), to enjoy life.

He Liveth long who Liveth well.

Or,

It is not how long but how well we Live.

v. Life consists, &c. under Life.

As long Lives a merry heart as a sad.

We may very well add, and longer too. The meaning however of this proverb is that immoderate sorrow tends to no good end.

One may Live and Learn.

Or,

We are never too old to learn.

In the opinion of Starnes and Noyes, the 1736 edition is inferior to the 1730 original, largely because of its chattier and more diffuse delivery. 'The author has yielded to the cardinal temptations which have beset lexicographers all along; he has included too many oddities and he has drawn no clear or consistent distinction between the provinces and methods of the dictionary and the encyclopedia.' To which they add, 'Bailey was merely of his time, whereas in innumerable other respects he was very much in advance of it.'

The book was however a commercial success, which is what Bailey was interested in. In his day, the distinction between dictionary and encyclopaedia had hardly been mooted, leave alone drawn, and one may wonder whether the distinction is yet as clear as Starnes and Noyes might wish.

With Bailey an era closed: the period when a dictionary could grow simply by feeding on its predecessors and amending their 'mistakes'. The groundswell for a standardizing dictionary to legislate in terms of the best literary English was now too powerful to be denied. Samuel

Johnson's Plan for a Dictionary of the English Language (1747)

together with the less well-known notes in Benjamin Martin's Lingua Britannica Reformata (1749) ushered in the new era of method and citation.

1.6 Johnson: Consistencies and Contradictions

Johnson was not a schoolmaster; and from journalism and drama he turned for an important period in his life to professional lexicography. His Plan and his subsequent Dictionary are widely considered to have 'raised English lexicography altogether to a higher level' (Murray: 42) and to have 'invested the calling with lasting dignity' (SN:196). Greene (1970:181) observes: 'Johnson's Dictionary is a very great and serious achievement in the history of the study of the English language. It is not merely a curious whim of a quaint eccentric, but a most important landmark in the development of English, from a set of unimportant local dialects spoken by a small group of islanders on the fringe of civilization, to a great world language.'

Johnson was invited to compile the long-awaited answer to the French Academy by a group of booksellers in London, and not by representatives of the elite. He was paid £1,575 to cover three years' work, and was not set at the head of 40 men over 40 years. Johnson was well acquainted with the current trends in language study and debate, and appears not to have been in the least daunted at setting out with a handful of mainly Scottish clerks to emulate the French. His work however took four years longer than he had planned. The labour in Gough Square, working from an interleaved folio edition of Bailey and passing marked books to his helpers is now legendary, but from it emerge

three cardinal points that affect much of subsequent dictionary-making:

- 1 it would be teamwork rather than the solitary effort of one man
- 2 dependence on previous compilers would be secondary to study of current 'best' usage
- 3 such usage would be presented citationally in as many assigned quotations as space permitted.

In Johnson's musty garret the great co-operative ventures of Oxford and Merriam-Webster had their beginning.

In his Plan he presented his aim of 'fixing' the language, in the Stoic belief that 'all change is of itself evil' and 'ought not to be hazarded but for evident advantage'. This view was echoed and enlarged by his patron Lord Chesterfield who, in surrendering his independence as 'a free-born British subject' to the new arbiter of the language, observed:

'I cannot help thinking it a sort of disgrace to our nation, that hitherto we have had no such standard of our language (as the French); our dictionaries at present being more properly what our neighbours the Dutch and the Germans call theirs, WORD-BOOKS, than dictionaries in the superior sense of that title. All words, good and bad, are there jumbled indiscriminately together, insomuch that the injudicious reader may speak, and write as inelegantly, improperly, and vulgarly as he pleases, by and with the authority of one or other of our WORD-BOOKS.

'It must be owned that our language is at present in a state of anarchy; and hitherto, perhaps, it may not have been the worse for it. During our free and open trade, many words and expressions have been imported, adopted and naturalized from other languages, which have greatly enriched our own. Let it still preserve what real strength and beauty it may have borrowed from others, but let it not, like the

Tarpeian maid, be overwhelmed and crushed by unnecessary foreign ornaments. The time for discrimination seems to be now come. Toleration, adoption and naturalization have run their lengths. Good order and authority are now necessary.'

Johnson set to work.

In the Preface to A Dictionary of the English Language (1755), he begins with the Stoic line that the language had been 'neglected', 'suffered to spread, under the direction of chance, into wild exuberance; resigned to the tyranny of time and fashion; and exposed to the corruptions of ignorance and caprices of innovation'. The only optimum in his observations is an admission that every increase of knowledge will demand new words and the tropes of poetry lead to new extensions of meaning and use. Such attitudes did not, however, survive the actual work. His hope to catch, fix and forever protect the language was abandoned in the light of eight years' experience. 'Fixing' became 'embalming', and 'to enchain syllables, and to lash the wind, are equally the undertakings of pride.' The pessimism at the end of his gargantuan work is deep: 'If the changes that we fear be thus irresistible, what remains but to acquiesce with silence, as in the other insurmountable distresses of humanity? It remains that we retard what we cannot repel, that we palliate what we cannot cure.' This is the only recourse of the Stoic, because 'tongues, like governments, have a natural tendency to degeneration.'

Intervention could only be a delaying action, never outright success. His pessimism contrasts strongly with the optimism of the hard-word compilers, who believed they were doing a service by their

kind of intervention. It is typical of the social potency of dictionaries, however, that Johnson has been regarded as successful in his aim: he might not have fixed the language, but he had the right idea, and slowed the rot. Greene (1970:178), puts this point of view when he says: 'Johnson, however, thinks stability in language is desirable, and he of course is perfectly right... English was rapidly becoming a world language; and in such a situation lack of stability and standardization must cause much greater inconvenience to communication than it would when English was merely the language of a handful of people in a small island off the western coast of Europe. Indeed, since the time of the publication of Johnson's Dictionary, the rate of linguistic change in written English, at least, has been notably retarded; and with modern methods of communication... the tendency of the language has been to standardization.'

It is difficult to be certain that Johnson appreciated the world-language role which English was just beginning to acquire in his day. Again, he might not have equated fixing the language with any tendency in the language towards being fixed. These are not the same thing, and Johnson's urge to fix the language arose from a fear that there was no such tendency. If there is indeed a tendency in language (especially in a literate society) towards being fixed, then Johnson need not have been so thoroughly pessimistic, but we unfortunately have no clear proof of it. Greene's remarks serve to indicate, however, that there is still today a strong current of approval for the policy of retardation, for the lexicographer who stands at the city gate and forbids entry to the Goths and Vandals. In fact, the great controversy over Webster's Third International Dictionary in the early 1960s is proof positive that Johnson still has his followers.

1.7 Orthography and Orthoepey

In an age which deemed standardization and correctness necessities after the profusion of the Renaissance, spelling and pronunciation were matters of great importance. In consequence, 18th-century lexicography was greatly exercised over orthography (the right way to spell) and orthoepey (the right way to speak).

Efforts towards a standardized spelling had been proceeding for decades, from incipient Tudor beginnings, and every dictionary (copying diligently from its predecessors) encouraged increasing standardization. Modern commentators often assume that confusion existed during the 17th and 18th centuries not only over spelling, but about the relation between the spoken sound and the written letter. This confusion was not however as abysmal as we tend to believe (cf Abercrombie on letters, 1949 (1965:76ff)). There was however a general optimism that just as the 26-letter alphabet could be used to everyone's advantage on paper, so refined people could harmonize their pronunciation, and compilers could appropriately indicate the 'best' pronunciation in their wordbooks. Just how to decide on the best pronunciation was not however clear. Some suggested following the House of Lords, others a spelling pronunciation accompanied by elocution.

The pioneers of orthoepey were in the main not Englishmen. The impulse appears rather to have come from the Irish and the Scots and to have linked speech to writing as much as to any prestige accent-form. Mathews comments as follows on the matter (31): 'Sheridan, though an Irishman, felt himself capable of grappling with the difficult questions of English pronunciation'. Mathews may have been ironic in this remark: there is no necessary reason why an Irishman should hold

back from such an undertaking, unless one has a belief that Irish speakers of English are in some sense defective. The example quoted by Mathews however indicates that Sheridan followed his own Irish norm in offering a standard -- a kind of bias not unusual in orthoepists.

Between 1769 and 1791 four dictionaries of orthoepy appeared, from James Buchanan, William Kenrick, Thomas Sheridan and John Walker. These developed the earlier but still commercially successful works by Thomas Dyche: A Guide to the English Tongue (1709) and The Spelling Dictionary (1723). Dyche placed an accent mark after the accented syllable, as in di'mner and di'ocese. William Kenrick (A New Dictionary, 1773) inserted the accent and also marked the separate syllables with numbers referring the reader to a table of pronunciation. John Walker (Critical Pronouncing Dictionary and Expositor of the English Language, 1791) used small superscript numbers indicating the 'powers' of the various written vowels, and this method proved so popular that a number of books appeared which claimed to blend the definitions of Johnson with the pronunciations of Walker.

A wide range of supplementary orthographies has been employed since then to help dictionary-users with pronunciation and accentuation. Discussion of this plethora would take up too much space here (cf Mathews, 76ff, for detailed examples), but the development of the purely orthoepic dictionary can be said to have faltered until the development of phonetic transcription. In 1886, for example, the International Phonetic Association was formed in France by a group of language teachers. Systems of diacritical marks and re-spelling to indicate phonetic values are still popular, however, in commercial dictionaries, and it is mainly in teaching English to foreign learners that phonetic transcription has been used.

Orthoepic entries in the dictionaries of the 18th century may have had a profound and lasting impact on speakers of English, especially in America. Read (1973:71) quotes to show that in the early 19th century, the dictionary of the Scotsman William Perry had had considerable standardizing effects throughout the United States. In 1828 one John Pickering observed that 'the Scotch dictionary of Perry' had had more effect on New England speech than any other, and that 'where we differ from the English, particularly in some of the vowels, it will be found that we agree with the Scotch.'

This would suggest that Perry had followed Sheridan in using his own accent as the most convenient model, a procedure followed in more recent times in England by Daniel Jones, in his An English Pronouncing Dictionary (1917). Gimson notes that in this work Jones 'shows only one type of British English pronunciation -- basically his own -- but indicates in the case of many entries a great variety of forms' (1973:115). Jones uses the symbols of the IPA, and avowedly records a specific accent called Received Pronunciation (RP) or Received Standard, which he accepts as linked with both region and class. It is 'the pronunciation used by a considerable number of typical Southern English people in ordinary conversation,' people 'who have been educated at the public schools'. In this instance, public refers to the system of private (largely boarding) schools in England. Jones does not explicitly advocate it as the 'best' type of English, but instead notes: 'The book is a record of facts, not of theories or personal preferences. No attempt is made to decide how people ought to pronounce; all that the dictionary aims at doing is to give a faithful record of the manner in which certain people do pronounce.'

Jones therefore disclaims authority, but since 1917 his book has been regarded in English-teaching circles as a prime authority. The orthoepic tradition predisposes people to take such works as authoritative, whatever their prefaces may say, and the selection of the accent of one class is, like the choices of Sheridan and Perry, implicitly a recommendation of the worthwhileness of that accent. It is prestigiously associated with royalty (as the King's or Queen's English), with an upper-class minority possessing significant social power, and indeed during the 1930s at least was openly the sole accent acceptable for BBC announcers operating from London (cf Lloyd-James, 1935; Abercrombie, 1951 (1965:10ff)). The accent is distinguished by diphthongization and a systemic linking 'r' and is considered to have the advantage of being very widely understood. At the same time, however, the Jones model is not now taken as a norm for speakers of RP, and is little other than a convenient device for teaching one English accent-type to foreigners (cf Gimson, 116).

RP poses the problem of any individual's or group's accent being elevated as a norm towards which others can aspire (whether that norm be Irish, Scottish or Southern English). It arises from the assumption that speech harmonization is possible or desirable, and since the 18th century has been linked to the idea of orthography: spelling standardization has been largely achieved, so speech standardization can also be achieved. The two are not however commensurate, since writing uses only 26 clearcut symbols and speech demands complex actions dependent on the articulatory set of mouth and jaw. There is no evidence that speech harmonization in the English-using world is an achievable aim, or that a specific model for foreign learners aids their ultimate intelligibility. It is questionable therefore whether *de facto* orthoepic works like Jones's dictionary (excellent as phonetic records)

should be offered as authorities. Lloyd-James, although in the 1930s he perpetuated the use of RP in the BBC, himself questioned the rationale of doing so: 'The dual education system of our country, State and private schools, reacts upon our ideas of standard pronunciation, and the steady growth of the State secondary school must cause us to think twice before we accept unquestioningly the traditional definitions of Standard English' (164).

Dictionary-compilers are still faced with fundamental problems, such as

- 1 whether to include pronunciation at all
- 2 whether to include a set of important alternative pronunciations
- 3 whether only to indicate the pronunciation of some rather than all words
- 4 whether to use diacritical marks, re-spelling or IPA symbols or some other system like Pitman's Initial Teaching Alphabet
- 5 how to mark accentuation or word-stress.

We have seen that to some extent the Jones model is a convenient fiction, and there is always a possibility that for World English in future another convenient fiction may arise, midwived by a council of phoneticians and lexicographers. It could have some such name as 'World Standard' and judiciously balance the contending forces. It might have an influence over general international English, for purposes of the widest intelligibility among the English-using nations, and could even proceed with the healthy proviso that it implies nothing pejorative about any accent anywhere.

1.8 The Encyclopaedia

In defining words, the compilers of dictionaries through the 15th to the 18th centuries teetered constantly on the edge of 'general knowledge'.

Various compilers have felt the urge to capture all knowledge in a single net (witness St Isidore himself), and this urge appears to have originated in the Greek conception of education. For them, the arts and sciences belonged in the enkuklios paideia, the 'encircling training' or general education of each generation. The Romans Quintilian and Pliny favoured an in-gathering of such necessary knowledge, and Renaissance scholars were not averse to collections of various kinds. It was left to the rationalists of the 18th century however to begin the work seriously.

The first wordbook in English with this kind of title was Ephraim Chambers's Cyclopaedia, or General Dictionary of Arts and Sciences (1728). This work developed the important, well-illustrated Lexicon Technicum of John Harris (1704), and is among the first attempts to provide special word-lists and definitions for the sciences. The main inspiration for future work was however French, in the Encyclopedie ou dictionnaire raisonne des sciences, des arts et des metiers (1751-72). As with the legislative dictionary, the French set the pace and the English were hesitant to follow. It should be noted however that both in Britain and France 'encyclopaedia' and 'dictionary' were synonymous at that time.

The Scots pioneered the English-language encyclopaedia. It was in Edinburgh that a 'Society of Gentlemen' published (1768-1771) the three volumes of the first Encyclopaedia Britannica, covering the arts and sciences with profuse notes and illustrations, but not at that time including biography and history.

The label 'Britannica' is significant. According to the present-day American publishers of this encyclopaedia, the title always referred

to the whole English-using world, the area encompassed by the British Empire of the time. We have seen also that there was a tradition in the Latin naming of wordbooks for 'Britannicus' to appear as often as 'Anglicanus' as the name for the language. The last two of this kind of title were Kersey's Dictionarium Anglo-Britannicum (1708) and Bailey's Dictionarium Britannicum (1730). It is reasonable to suppose that these titles were influenced by the increasingly close association of the two kingdoms of England and Scotland after the union of the crowns (1603) and the union of Parliaments (1707). Johnson chose the title 'English Language' in 1755, but there were precedents for a possible 'British Language', on the analogies of these Latin titles. Johnson was not however known for his willingness to accommodate the Scots.

These competing titles point to an important and often forgotten historical fact. As the name for a language, 'English' antedates the name for a political entity called 'England'. In its earliest state, the Anglo-Saxon complex was spoken on both sides of the North Sea; later it spread through the eastern districts of the island which was later divided into the polities of Scotland, England and Wales, competing with various Celtic tongues. It was never therefore the national possession of any people, whether or not a national name happened to be identical to the name of the language. To cope with the amalgamation of North and South the old Celtic name 'Britain' was resurrected, but not applied consistently to the common language. Indeed to call 'English' British would have been a historical and linguistic misnomer. It seems reasonable however that an encyclopaedia dealing with general knowledge should retain the Latin cover-term.

Some weight was in due course added to the decision of the Scots encyclopaedists to favour a term other than 'Anglicana'. Their three volumes were not long in circulation before the outbreak of the American War of Independence (1776).

1.9 Webster: Consistencies and Contradictions

The extension of English through colonization posed a new problem, quite distinct from the infusion of Neo-Latin or the need for a literary standard. Communities separated by thousands of miles, living in very different environments and with divergent social ideals can hardly keep their language undifferentiated, and if they have fought each other, can hardly be expected to want to. The 'mother country' would label special expressions from the colonies and ex-colonies 'barbarous', and the term 'Americanism' with its pejorative ring would be added to 'Scotticism', 'Irishism' and the like. The Americans in turn would assert that their usage was every bit as good as in a far-off dictatorial island -- and indeed better.

The optimism of the new land prevailed. Mathews quotes an anonymous 'American', writing before the break, in 1774, that 'the English language has been greatly improved in Britain within a century, but its highest perfection, with every other branch of human knowledge, is perhaps reserved for this Land of light and freedom' (36). Suggestions were later made that an American Academy be set up to supervise this vast amelioration of the English language, but as in Britain the achievement was left to individuals.

Noah Webster was not the first compiler of American-English word-books, but he is the first serious innovator among American lexicographers.

He holds the same position in the American mythology of letters that Johnson holds in Britain and in his long life (1758-1843) spans the formative period of American history. Webster was not an admirer of Johnson, but they had much in common: they set out to tame the lexicon, standardize what might be standardized, and provide the right citations for the best use and appreciation of the language. They shared a faith in good writers, Webster quoting Johnson ('The chief glory of a nation arises from its authors') as justification for adding quotations from the American Founding Fathers to pre-existing English-national writers. In this he did more than Johnson to internationalize the language.

Both men were Stoic in outlook; we have seen Johnson's pessimism and his final aim of retarding corruption. Webster had a particularly pure Stoic ideal: 'to ascertain the true principles of the language, in its orthography and structure; to purify it from some palpable errors, and reduce the number of its anomalies,' and so to proceed to a standardization fit for the future 'three hundred millions of people' whom he confidently predicted would inhabit the United States (1828). He differed from Johnson on two counts:

- 1 he included most lexicographers as among the corrupting influence on language
- 2 was curiously shot through with optimism.

At one and the same time he wanted to stave off 'the mischievous influence of ... that dabbling spirit of innovation which is perpetually disturbing settled usage' and looked forward to great advances in his pioneering homeland.

Johnson was not Webster's idea of a model, and he contributed to his own unpopularity at home and in Britain by strongly criticizing both Johnson and Johnson-worship. In this he saw himself as a latter-

day Copernicus, a publisher of truth (1807:125), taking heart from the opinion of John Horne Tooke, a British etymologist, that Johnson's works were 'most truly contemptible performances; and a reproach to the learning and industry of a nation, which could receive them with the slightest approbation'.

His own conclusion was simple: 'I am prepared, by a minute examination of this subject, to affirm, that not a single page of Johnson's Dictionary is correct' (140). And apart from that he castigated his predecessor on seven points:

- 1 the insertion of Latin barbarisms on the sole authority of earlier compilers like Bailey and Phillips
- 2 injudicious choice of authorities to quote, as for example the Latin-oriented Thomas Browne
- 3 the inclusion of low and vulgar words
- 4 the lack of satisfactory sense discriminations in defining words
- 5 failure to distinguish near-synonyms properly
- 6 quotations which often fail to illustrate meaning and in the case of the common words of English are redundant anyway
- 7 appallingly inaccurate etymologies

Lexicographers have often put out such criticisms, only in time to be hoist with their own petard. In the question of etymology alone - on which Webster fulminates with special heat - there is evidence in plenty that Johnson was inaccurate and lacking in knowledge.

Unfortunately, Murray (43) and Mathews (41f) both upbraid Webster for his unbridled etymological fantasies and tendency to invent. Like Isidore of Seville, Webster believed in the Tower of Babel and assumed that a quick mind with the right theory could see affinities between words far separated in time and place, without needing to check the available records.

A schoolmaster by profession, Noah Webster began his publishing career with the famous Elementary Spelling Book (1783). In 1806 there followed his Compendious Dictionary of the English Language and in 1828 (when he was 70 years old) his 70,000-entry An American Dictionary of the English Language came out. Definition, citation, orthography and etymology were his mainstays, together with encyclopaedic elements such as all the post offices in the United States, weights and measures, currencies and chronology. His spelling plan has been particularly subject to scrutiny and criticism, and rested on the principles of reversion to earlier models and analogical resemblances. Color and labor are nearer to Latin and hence more desirable than colour and labour. Some spelling reforms survived; some did not. Some stayed resolutely at home; others conquered the old country. His physic and logic are now universally preferred to the older British forms physick and logick. Webster added many words to the lists of what were acceptable, including such controversial items as applicant, departmental, presidential, advisory, ascertainable and subsidize, very much a part of today's vocabulary and apparently as barbarously Latinate as anything in Johnson.

Webster had his admirers, but was not generally popular with his fellow Americans in his lifetime. He also had a rival in Joseph Worcester, who had greater personal success at the time but who has been eclipsed by Webster's afterglow. Worcester began his career as a careful reviser of Johnson and abridger of Webster, but branched out into a series of three works: A Comprehensive Pronouncing and Explanatory Dictionary of the English Language (1830); A Universal and Critical Dictionary of the English Language (1846); and A Dictionary of the English Language (1860). This last covered 104,000

entries, the largest collection to date. Synonyms were given attention and illustrations so well integrated into the text that Worcester is sometimes given the credit of introducing illustration as a technique.

Whitehall (1958 (1971:162)) observes: 'Temperamentally the flamboyant Noah Webster and the cautious Joseph Worcester were natural rivals', but their rivalry was minimal compared with the commercial war between their respective publishers. Worcester was conservative and a recorder; Webster was radical, innovative and iconoclastic. These contrasts however dulled in a war of the dictionaries where rival publishers plagiarized, pushed and manoeuvred to drive each other off the market. In the United States dictionaries graduated from good steady business to very big business. Immigration, expansion and a public-schools programme demanded wordbooks. Flaws of style and method were eliminated in the fierce competition and Whitehall considers that the rivalry had the side-effect of promoting better lexicography. Some time after Noah's death, the Webster dictionaries, published by Merriam, won the field, but in the process had so synthesized the work of the two rivals that it is probably best to link them posthumously as the Webster-Worcester tradition.

1.10 The Dictionary on Historical Principles

Back across the Atlantic (but aware of the American developments) lexicography entered a new phase. The hard-word dictionary had left its legacy on how people thought about wordbooks, as had the standard universal dictionary - but in the 19th century in Britain concern was developing over a new historical view, some expression of unfolding change (and perhaps improvement in keeping with the scientific spirit of progress).

Two men stand out in this approach: Charles Richardson in his New Dictionary of the English Language (1836), and Richard Chenevix Trench (1857) in his two addresses to the Philological Society On Some Deficiencies in Our English Dictionaries.

Richardson was impressed both by the copious quotations in Johnson and by the etymological ideas of Horne Tooke, and considered that quotation alone ought to provide meaning. Not only that; graduated quotation through time would show change and development. He did not dispense entirely with short definitive statements, but sought to make them dependent on the quotations rather than have quotations simply to buttress definitions. Murray (44) considers this approach excellent in theory, but vitiated by the practical problems of space: 'Quotations will tell the full meaning of a word, if one has enough of them; but it takes a great many to be enough, and it takes a reader a long time to read and weigh all the quotations, and to deduce from them the meanings which might be put before him in a line or two.' Murray, with this proviso, remained impressed by the mass and exactness of Richardson's work.

The clergyman Trench was largely in agreement with Richardson. The Philological Society had been formed in 1842 to investigate the structure, affinities and history of the English Language and had begun to collect words not found in Johnson or Richardson, intending to supplement them in some way. Trench saw the Society as the vehicle however of a more original undertaking and pointed out to its members (of Trench, reprinted in Sledd and Ebbitt, 1962:36) that no dictionary to date had adequately handled the following seven points:

- 1 obsolete words
- 2 derivational families
- 3 accurate citation, with dates
- 4 all important senses of a word
- 5 the accurate distinguishing of synonyms
- 6 sufficient coverage of the available literature
- 7 elimination of redundant material

Removing the more eccentric comments in Webster, we find the American and the Englishman very largely agreed. Trench however differed from Webster in honouring the achievements of Johnson and the others, and urged the Society to undertake its own comprehensive dictionary on historical principles.

In making this recommendation, Trench defined the word dictionary in a way which detached it from Chesterfield's dictum about social desirability. Logically, for Trench, a dictionary is an inventory of the language, and 'it is no task of the maker of it to select the good words of a language... The business which he has undertaken is to collect and arrange all the words, whether good or bad, whether they do or do not commend themselves to his judgment... He is an historian of it, not a critic.' Trench kept the literary orientation of lexicography, but dismissed its legislative function. The door must henceforward be kept open for all the waifs, strays, monsters, misfits and casualties of the centuries, as well as those deemed fit and worthy.

Richardson and Trench helped to inspire a new philological approach to the compilation of dictionaries, but the influences at work were not solely within Britain or English language studies alone. For some time

German scholars such as Schneider and Passow had been interested in an adequate record of the vocabulary of classical Greek. Schneider's Lexicon appeared in 1797-8 and Passow's in 1819. Passow had laid down in his Essay in 1812 (Zweck, Anlage, und Ergänzung griechischer Wörterbücher) new canons of lexicography, which included the requirement that citations be chronologically arranged so that the history and uses of each word would be clearly seen. The Greek-English Lexicon of Henry G. Liddell and Robert Scott () was based on such principles. X

Such were the influences at work when the members of the Philological Society concurred with Dr. Trench, and the 70-year story of the Society's great dictionary began.

1.11 Murray's Dictionary

It is impossible to do justice in a section of a chapter to the mammoth project which the Society undertook, or to its ramifications. Here, one can only glance at the achievements of its remarkable editors, and seek some assessment of the methodology which emerged.

When the work began no one could have conceived the shape it would take, or the time necessary for some kind of completion. Perhaps if it had been possible to foresee the effort involved, no one would have come to Dr. Trench's assistance. In the event, however, four men pre-eminently carried the work from its amorphous start to its massive ordered finish, a triumph of 19th-century optimism. These men were:

- 1 Robert Coleridge, the first, youngest and shortest-lived of the editors, who supervised two committees from 1858 to his death in 1861. These committees dealt with literary and historical sources on

the one hand, and etymology (now a very important matter) on the other. Coleridge enthusiastically engaged volunteer helpers in considerable numbers on both sides of the Atlantic.

2 Frederick J Furnivall concerned himself over 20 years with ensuring that such volunteers had accurate texts to work from and a coherent reading plan to follow. To this end he founded the Early English Text Society in 1864 and the Chaucer Society in 1868, and originated the plan of organizing quotations under the numbered senses of a word. In 1879 when he handed on the work, $1\frac{3}{4}$ tons of material had accumulated in his home. Furnivall negotiated with the Delegates of the Clarendon Press, Oxford, and obtained from them a contract promising to print and publish the dictionary, and also to support a full-time editor, assistant and two clerks. With that success, compilation proper could begin.

3 James A Murray benefitted from Furnivall's agreement with Oxford.

He worked on the dictionary for the remainder of his life and was responsible for alphabetizing the mass of material passed on to him, which was housed in its own iron building with 1,029 pigeon-holes for slips sent in by some 2,000 readers. His first section, A-Byz, came out in 1888 and his preface to it remains a classic of its kind.

4 William A Craigie first worked under Murray, editing sections of the work along with Henry Bradley and Charles T Onions. Working mainly in the Old Ashmolean in Oxford, he brought the project to successful completion in 1928.

The great dictionary was started off by two English editors and finished by two Scots editors and has not been easy to name. According to its first title, it is the New English Dictionary (NED), but it has also been known descriptively as the Historical English Dictionary and as Murray's Dictionary, after the principal editor. Since 1933 however

the Clarendon Press has referred to it as the Oxford English Dictionary and this last seems to be establishing itself as the commonest title. Because of the association with Oxford through publication and geography, the title is suitable, but it does have a flaw in that the originating function of the Philological Society has been largely forgotten, and casual association can exist between an Oxford dictionary and the prestige of an Oxford English, where no such association was originally intended.

This is the first dictionary to which staggering statistics can be attached, establishing a fashion of 20th-century dictionary-description. It took 70 years of compilation, had 5,000,000 quotations sent in by 2,000 readers, 1,800,000 of these quotations eventually being used, took forty days for the organizing of the one word set, and contains 414,825 entries. Whitehall (161) probably speaks for most commentators when, as an American lexicographer, he says: 'No other language has ever been recorded on anything approaching this scale, and no dictionary of English since the New English Dictionary was completed has failed to reveal a profound debt to this monumental work.' Like Whitehall, I shall refer to this work as the New English Dictionary (NED).

The aim of the NED is best summed up in Murray's own words in 1900: 'not merely to record every word that has been used in the language for the last 800 years, with its written form and signification, and the pronunciation of the current words, but to furnish a biography of each word, giving as nearly as possible the date of its birth or first known appearance, and, in the case of an obsolete word or sense, of its last appearance, the source from which it was actually derived, the form and sense in which it entered the language or is first found

in it, and the successive changes of form and developments of sense which it has since undergone' (47).

At that time Victorian optimism dominated, and Murray said: 'Be the speed (of compilation) what it may, however, there is the consideration that the work thus done is done once for all; the structure now reared will have to be added to, continued and extended with time, but it will remain, it is believed, the great body of fact on which all future work will be built. It is never possible to forecast the needs and notions of those who shall come after us; but with our present knowledge it is not easy to conceive what new feature can now be added to English Lexicography. At any rate, it can be maintained that in the Oxford Dictionary, permeated as it is through and through with the scientific method of the century, Lexicography has for the present reached its supreme development.' (49).

His claims are put with care and precision, but they are not humble claims. And yet as the compilers themselves moved through time with their mammoth task, the language was moving with them, changing, its fund of words growing, pressing forward to be registered. Thus, when the compilers reached 'S' in 1910, they were 26 years away from the compilation of the letter 'A', a period long enough to justify a revision among less epic dictionaries. And such was the recognition in the end of the slow power of time that, when the dictionary was completed in 1928, its first Supplement was immediately undertaken.

After Murray's death, Craigie touched upon the imperfections of the ultimate dictionary, proposing in a paper to the Society in 1919 that further work lay beyond: work on period dictionaries for Old

English and Middle English, and Early Modern and Late Modern, and on The Older Scottish Tongue, and American English and so on. With regard simply to Late Modern English, Craigie amended the historical principle with what seems a nod to De Saussure: 'The historical element diminishes, the practical side becomes more prominent. During the 18th century, the scientific terminology had been steadily increasing; in the 19th it assumes overwhelming proportions... To deal adequately with this demands a dictionary in which there will be little room left for the historical element, because all the available space will be required for the exposition of facts as they are, without regard to what they may have been in the past' (in Mathews, 72).

There was, in fact, still plenty to be done.

1.12 Conclusion

The Philological Society's Dictionary is a suitable point at which to close a chapter. It stands on a summit of achievement in the glossary tradition, with that tradition's emphasis on alphabetic or 'dictionary' order. Not only does it seek to catch all the words used in the last 800 years, but it is itself the culmination of 800 years of effort at explaining those words.

In this chapter we have seen that lexicography began more or less with the parallel glossary and vocabulary traditions of the Middle Ages. So far we have examined the developments in the glossary tradition, and have yet to look at its partner. In the glossary tradition there is an easily traceable line from interlinear glosses to collections that put Latin first, then English first, and finally that bring the Latin

into English in explanatory 'hard-word' dictionaries. These wordbooks were responses to the lexical intrusion of Neo-Latin into English during and after the Renaissance and were an attempt to increase the number of people capable of appreciating writers who were bilingual in Latin and English.

The hard-word dictionary gave way to the legislative-literary dictionary, during a period when men sought to 'fix' their language in a refined and perfected condition. Johnson began his work with this aim in mind, but ended it pessimistically concluding that it would be enough simply to retard the corruptive changes at work inexorably throughout the language. From the efforts of the Kersey-Bailey-Johnson tradition, however, emerges a strong impetus towards 'standard' and 'universal' dictionaries, an impetus which underlies the 19th-century aim of achieving the truly comprehensive English dictionary.

Efforts were made on both sides of the Atlantic to get a really comprehensive wordbook, one in which 'all' the words of English could be set out with adequate citations and etymology. This development coincided with and received inspiration from the philological advances of the time, and encouraged a belief that the glossary tradition could encompass the task of accurate, detailed recording without legislative comment. The NED under Murray represents the flowering of this belief.

2 The Vocabulary Tradition in English Lexicography

2.1 Introduction

Wordbooks and the alphabetic ordering of words are so closely associated in the modern mind that alternatives receive little attention. The vocabulary (traditionally organized without recourse to the alphabet) is usually regarded by students of lexicography as an aberration, appearing here and there as a mediaeval list, a Tudor dictionary or a Victorian thesaurus. Historians of lexicography like Murray and Mathews have accorded vocabularies enough space to indicate that they affected the development of dictionaries, but have not generally seen them as a tradition in their own right.

I hope to demonstrate in this chapter that the vocabulary has a strong tradition of its own, and that many important innovations in general lexicography and the study of word relationships belong to that tradition rather than to the glossary tradition.

Historians for the vocabularies are scarce. Starnes (in SN, 1946:197ff) provides a detailed appendix on Mediaeval and Renaissance vocabularies, while in the introduction to the Merriam-Webster Dictionary of Synonyms there is an essay by Rose F Egan entitled Survey of the History of English Synonymy (1942) which takes up largely where Starnes leaves off. We now have a detailed biography of Peter Mark Roget (Emblen, 1970), to supplement and expand information in the introduction to the Thesaurus. From these and other sources, and from the works themselves, one can piece together something of the vocabulary tradition, the overshadowed side of lexicography.

2.2 Models of Cosmic Order

Special vocabularies date at least from Aelfric, Abbot of Eynsham, in the 10th century. They were compiled, not to help students understand occasional words in texts, but to provide them with overall vocabularies of Latin for their various fields of interest. Aelfric's lists had Latin first and Anglo-Saxon second, and covered thirty topics:

1 agricultural tools	16 birds
2 ecclesiastical affairs	17 herbs
3 officials of the church	18 trees
4 officials of the state	19 colours
5 Roman law	20 clothes
6 man	21 games
7 his kindred	22 amusements
8 the parts of the body	23 weapons
9 diseases	24 heaven
10 the house	25 earth
11 its parts	26 sun
12 its contents	27 moon
13 food	28 angels
14 drink	29 archangels
15 beasts	30 ships and their parts

The original Aelfric list appears to have been compiled with only the roughest attempt at a system. Material in the 11th century, though without the topic headings in the original, appears to run in a more systematic fashion, and Starnes regards the 11th-century version as a conventional order of topics. Formed into a similar list, these topics are:

- 1 God, heaven, angels, archangels, sun, moon, earth, sea
- 2 man, woman, the parts of the body
- 3 terms of consanguinity, professional and trades people, artisans
- 4 diseases
- 5 abstract terms
- 6 times of year, day, seasons, weather
- 7 colours
- 8 birds
- 9 fishes
- 10 beasts
- 11 herbs
- 12 trees
- 13 house furnishings
- 14 kitchen and cooking utensils
- 15 weapons
- 16 parts of the city
- 17 metals and precious stones
- 18 general -- abstract and concrete

The headings serve to cut up the continuum of life into manageable sections, arrived at by subjective decision but expressing more than just individual preference. The list gives us some idea of what was centrally significant to Aelfric and his contemporaries. As a retrieval method such conventional topics were as convenient to use as the crude alphabetizations of the time, but as alphabetization advanced to the third and fourth letter it proved a safer and more popular tool than these arbitrary subdivisions of reality.

Like the glossary, the vocabulary benefited from the development of printing in the 15th century, and was as commercially successful.

John Stanbridge produced his Vocabula in 1496 and his Vulgaria in 1508. These contained the conventional-topic list, and were followed in 1519 by William Horman's Vulgaria, which contained the innovation of Latin and English parallel sentences illustrating words at work. The sentences appeared under abstract headings like De Pietate ('on piety') and De Animis Bonis et Malis ('on good and evil spirits'). Robert Whittinton's Vulgaria in 1520 synthesized the two, having the conventional topics and the illustrative sentences. Like the glossaries of the time, these wordbooks are forgotten now, but were commercial and educational successes in their day. They did not seek to explain hard words, as can be seen from the title vulgaria ('common things'). That they offered no lasting competition to books which did explain the hard words of Latin or English seems clear from the failure of vulgaria to transfer into English as 'vulgary', while dictionary transferred as 'dictionary'. Indeed, even the term 'vocabulary' as a name for a wordbook has only established itself sporadically.

The next significant book in the tradition was in fact called a 'dictionarie', and put the English words first. It is John Withals's A Shorte Dictionarie for Yonge Begynners (1553). It is non-alphabetic, has the conventional-topic list and is philosophically-inclined. Starnes notes (SN:203): 'Beginning with the aether, for example, he mentions the twelve signs of the Zodiac; he continues with the seven planets, the divisions of time, the seasons, the four elements of fire, air, water and earth, and the common words and ideas associated with each. So throughout, the author sought to suggest by arrangement a close relationship of words with groups as well as connections between groups themselves. Apparently Withals shared the belief of his time in a definite cosmic order.'

Withals's style of presentation of the word groups within this cosmic framework is simple but effective. A yonge begynner would not fare too badly in the following:

A Ship with other water vessels etc

A ship, Navis, rates
 He that maketh the ship, Naupegus, gi
 The keale or bottome of a ship, Carina, ae
 Hardie shippes, Audaces carinae
 The keale of a ship pitched, Uncta carina
 The fore part, Prora, ae
 A great ship, Trieris, ris
 The snoute afore, Rostrum navis
 The hinder part of a ship, Puppis, pis

Withals had garnered his material from three sources:

- 1 the other schoolmasters, Stanbridge, Herman and Whittinton, who had immediately preceded him in the tradition
- 2 the dictionary of Sir Thomas Elyot, a bilingual glossary mentioned in chapter one
- 3 the conceptually-based work of a Belgian, Joannes Paludenus (1549).

In his presentation the vague approximations of the early Aelfric material shape themselves into a special Renaissance world-view. The compilation was a success, and was reissued regularly up to 1634.

It was a religious educator who developed the approach further, William Bathe, an Irish Jesuit working in Spain who in 1611 brought out a Latin-to-Spanish manual of a similar type. A Latin-to-English equivalent appeared in London in 1615, entitled Janua Linguarum ('the gate of tongues'). It also was a success, surviving through nine editions to 1645. It consisted of some 5,000 items classified in twelve conventional sections with the words fitted into 1,200 explanatory sentences. Bathe was rigorous in keeping to a general plan and in teaching morality through lexis, and stirred the interest of another religious educator, John Amos Komensky in Moravia. Wishing to take

Bathe even farther, Komensky, better known as Comenius, brought out in 1631 a work called Janua Linguarum Reserata ('the gate of tongues unlocked'). It was first printed in Latin and German, but in the same year appeared in England in English, French and Latin. In the course of a few years it spread with improvements and variations to most of the Western European languages and had a great influence in language teaching. Its content and method remained essentially that of Bathe with minor changes such as the gradation of sentences according to difficulty. It had a hundred chapter headings and 1,058 sentences, and an analysis of the topics presented helps to show how the Mediaeval moved through the Renaissance towards the encyclopaedic interests of the Enlightenment. The numbers in the following list do not reflect Comenius's chapters, but indicate that 19 main topics can be abstracted from the whole:

- 1 the origin of the world
- 2 the elements, the firmament, fire and meteors
- 3 waters, earths, stones, metals
- 4 trees, fruits, herbs, shrubs
- 5 animals
- 6 man and his body, with external and internal members
- 7 the qualities or accidents of the body
- 8 diseases, ulcers and wounds
- 9 the senses, external and internal
- 10 mind, the will and the affections
- 11 the mechanic arts
- 12 the house and its parts
- 13 marriage and the family
- 14 civic and state economy
- 15 grammar, dialectic, rhetoric, arithmetic, geometry

- 16 ethics
- 17 games
- 18 death and burial
- 19 Providence, God and the angels

The cosmic ordering bears a strong family resemblance to Aelfric, but the dominance of nature gives way to more interest in human activities. The philosophical scheme dictates the nature of the word groups, but like earlier divisions does not offer a systematic procedure for sub-division and subclassification. The work was successful, but Comenius was not himself satisfied with it, and in 1657 supplemented it with a rather different book, the Orbis Sensualium Pictus ('the world of sensible things drawn'), a pictorial wordbook for language-teaching purposes. Comenius considered that clear sense impressions are the foundation of all learning, and so emphasized the need for visual assistance. The text contains 151 chapters, ranging from the Creation to the Last Judgment, each chapter being illustrated with a composite cut or engraving and each item in the illustration having its number. The sentences in the sections attached to the pictures have their numbers too, corresponding with those in the pictures. Thus, a picture of a study with its numbered parts corresponds to a sentence: 'The Study, 1, is a place where a Student, 2, apart from Men, sitteth alone, addicted to his Studies, whilst he readeth Books, 3...'

Comenius here applies the correlational power of numbers as effectively as the glossarists had learned to apply the ordering power of the alphabet. This innovation of matching numbers has been systematically exploited ever since Comenius, in all manner of books, and the particular strength of the tradition in Germany is demonstrated in the Duden wordbooks.

The Duden pictorial wordbook came late into English (1936), and in 1960 Harrap issued a revised and up-dated version called The English Duden: A Pictorial Dictionary with English and German Indexes. The use of an index (an offshoot of glossarial alphabetization) supplements a pattern which is still essentially Comenius. The modern book handles 25,000 words through 368 illustrated sections. Under each detailed and numbered picture is an equivalently numbered list of the words and phrases (but not sentences) that people use to denote the various objects in the pictures. Clearly, such a work is restricted to picturable material and cannot cover subtle semantic inter-relationships among words, but its range is impressive. The picture-and-list sections belong in sequence within a classificatory system which is a linear descendant of the cosmic orderings of Aelfric and Comenius, but vastly expanded by means of detailed subdivision of main topic headings. The user can go from index to section and get what he wants, or can go from topic headings to subheadings to sections. We shall see in later developments in the vocabulary tradition that supplementation of conventional topics by means of an index seems inevitable, and should remember that however important the index appears in Duden today, it was absent in vocabularies earlier than Comenius, and is absent in Comenius himself.

Comenius offers two distinct works which indicate a splitting in the vocabulary tradition. Duden follows the Orbis, while other compilers continue the line presented in the Janua. It is worthwhile however at this stage to list the main topics and some subheadings in the modern English Duden to show how things have changed since Comenius, to see just what has happened to the western European world-view. The subdivision of some sections in the Duden is very detailed, and only a



glimpse of it can be given here. The material gathered under State and City here is not intended to be taken as complete:

- 1 atom, universe, earth
- 2 man and home
- 3 horticulture, agriculture, forestry
- 4 hunting and fishing
- 5 crafts and trade
- 6 industry
- 7 printing and allied trades
- 8 transport and communications
- 9 office, bank and stock exchange
- 10 state and city
 - 1 money (coins and notes)
 - 2 flags
 - 3 heraldry, crowns
 - 4 police
 - 5 administration of justice (in Germany)
 - 6 school
 - 7 university
 - etc
- 11 travel and recreation
- 12 sport
- 13 entertainment and music
- 14 science, religion and art
- 15 animals and plants

Duden's cosmic and human order stands far removed in style and in detail from the mediaeval simplicities. Atoms dominate, angels are part of church architecture, and God is off-listed. It is clear from the similarities however that methodology can survive radical adaptations in world-view, and Comenius remains one of the great originators in lexicography at large. It is appropriate therefore to place him at

the end in a Paludanus-Withals-Bathe-Comenius sub-tradition. Of him Starnes says (208): 'In brief, there were current before Comenius all the elements that are found in his Janua and Orbis -- the tested topics under which the words were grouped, the wide range, implying universality, the use of sentences to give coherence to word lists, and the arrangement in parallel columns with the vernacular first. But no other book, so far as I know, exhibited the particular combination of these elements that is found in the manuals of the Moravian bishop and educator.'

The best-known native English development of Comenius is James Greenwood's The London Vocabulary (about 1700). It places English first, has 33 chapters or topics arranged in parallel columns, the English words being numbered to relate to composite cuts which occupy about a third of each page. The wording is not unlike Withals:

In Government there are

A Law 1	Lex, egis f.
An example	Exemplum, i n.

In Law there are

A Judge 2	Judex, icis m. and f.
A Counsellor 3	Consultor, oris m.
A Witness 4	Testis, is m. and f.

After Greenwood, the pictorial tradition merges with the glossary tradition, in both dictionaries and encyclopaedias, and we should turn elsewhere to find developments in the non-alphabetic vocabulary.

The division of a vocabulary into organizational topics either impels the compiler towards philosophy, or arises from an inherent interest in it. Aelfric of course possessed the mediaeval world-view of a Catholic Christian abbot and the Elizabethan schoolmaster Withals by and large shared it. By the time of Comenius, clergymen

were blending their religious conception of the universe with a newer scientific view, an outcome which did not have immense repercussions in the lexicography of the period but, lying dormant, was to affect a later generation.

In 1664, the Royal Society created a committee whose brief would be, quite simply, the improvement of the English language for scientific and philosophical ends. A member of the society whose name is linked with that committee was Bishop John Wilkins, who in 1668 produced a detailed and recondite work entitled An Essay Towards a Real Character and a Philosophical Language. This work was not concerned with the Latin/English schoolroom problems of Withals, but rather with setting up a system of universal concepts, each to be matched with a symbol which would enable scientists, regardless of language, to communicate with each other on their own special plane of abstraction. In his Essay, Wilkins discusses 'universal philosophy', 'philosophical grammar' and 'a real character' (= a thing-related writing system) for a 'philosophical language'. The discussion is detailed and with regard to the shortcomings of language and writing systems, he observes: 'It cannot be denied, but that the variety of Letters is an appendix to the Curse of Babel, namely, the multitude and variety of languages.' Wilkins seeks to rise above the flux and failure of natural language and to do so proposes a scheme strongly resembling Aelfric, Withals and Comenius. In proposing it he talks however of the genus's (sic) or categories of nature, and lists them much as follows:

- 1 God
- 2 the elements and meteors, stones, metals
- 3 plants, herbs, flowers, shrubs, trees
- 4 animals, fishes, birds, beasts

- 5 parts of bodies
- 6 quantity, magnitude, space, measure
- 7 quality (of natural power, habit, manners, the senses, diseases)
- 8 action (spiritual, corporeal, in motion, in operations)
- 9 relations (in the family, regarding possessions, provisions)
- 10 public relations (civil, judiciary, naval, military, ecclesiastical)

This presentation is somewhat simpler than the original Wilkins and abandons some of the terms which are no longer immediately interpretable. One such is, for example, 'oeconomical' in the sense 'domestic, pertaining to the house or home'. The cosmic scheme is linked with an alphabetical dictionary appended to the main discussion, and in this dictionary or index certain devices are used to send the inquirer to specific subdivisions of Wilkins's scheme. For example, the entry planet carries the code W.II.3, which means that the inquirer should look for information about planet under the section W (= world), its second subdivision and the third stage in that subdivision. Similarly, posterity is coded as RO 1.1.0, which gives the appropriate place under RO (= Relation Oeconomical). All the information can be given either in English or in the supralinguistic philosophical 'real character' which Wilkins is actually advocating.

The book's very abstruseness militated against its success. It contained seeds of the future, however, and we shall see in due course how one of these seeds grew.

2.3 Discriminating Synonyms

In the hard-word dictionaries of the 17th century synonymy was the translation of Latinate words by means largely of words of Anglo-

Saxon provenance. In the development of ideas of a fixed and refined language the need not only for mutual explication of but also careful discrimination among similar words gained ground, and like many other features of lexicography was precipitated by developments in France.

In 1718 there appeared a compilation of French synonyms by the Abbe Gabriel Girard. Nothing comparable however appeared in English until 1766, when John Trusler published two small volumes entitled 'The Difference between Words, Esteemed Synonymous, in the English Language, and the Proper Choice of them determined together with so much of Abbe Girard's Treatise, on this Subject, as would agree, with our Mode of Expression.' Trusler found that most of the Abbe's work was agreeable, including his preface, whose translation appears intact before Trusler's own remarks. He refers to the great esteem in which the French work was held, and its many editions, taking these to speak for its usefulness. Following Girard, he takes the view that synonymy is seldom complete: 'We must be a little nice upon words, not imagining, that such, as are called synonymous, are really so, exactly uniform in their sense.'

Trusler is non-alphabetical, arranging his sets of words in numerical order, sometimes in apparent haphazard succession, sometimes with some suggestion of a semantic scheme. Compare for example:

248 beam, rafter
249 room, chamber, apartment
250 house, tenement
251 lodging, apartment

but

282 to lower, let down
283 plant, herb
284 to grow, increase
285 to rise, get up
286 copy, model

Since Trusler dispensed with the idea of an overall conceptual framework (unnecessary for a relatively small work on synonyms), the system of numbering is essentially unmotivated. There are 370 sets in all, mainly of pairs and triples, so that we can assume a total vocabulary of between 740 and 1,110 words. The index added to the second volume is a very practical retrieval aid, all the sets being adequately indexed. A set which has the order rafter, beam, for example, is listed as such in the index against the number 248. So also, however, is beam, rafter, so that if a word is discussed, the index will lead you to it. This device was not however followed by later synonymists.

Discrimination is by means of short essays, contrasting the uses of the two or more 'synonyms', a discursive style which is still dominant in dictionaries of synonyms and antonyms. It is well illustrated in a work which followed Trusler in 1794, the British Synonymy, which is important for several reasons. Firstly, it was compiled by a woman, Hester Piozzi or Thrale, a friend of Samuel Johnson. She is the first of several women who have found the synonymy an avenue into lexicographic work. Secondly, in her title, she was successful in using 'synonymy' in the sense of a book about synonyms, but unsuccessful in using 'British' as the name for the language of the United Kingdom. Thirdly, she deliberately went against the scholarly current by examining conversation rather than the works of acknowledged writers. She considered that men should be concerned with the compilation of dictionaries proper and the teaching of writing; a woman, however, 'at worst' is qualified 'through long practice - to direct the choice of phrases in familiar talk' (preface, ii). She even suggested that 'synonymy has more to do with elegance than truth' (in the sense of logic), an observation which suggests that she was one of the first sociolinguists.

Mrs Piozzi was accused in her time and later of flippancy and shallowness (cf Emblen, 1970:263ff). In the constricted world of the drawing-room it is not surprising that she felt her own work to be rather light, but Emblen's point that she had 'a delicate ear' for nuances is justified. Two examples show her method:

'ABANDON, FORSAKE, RELINQUISH, GIVE UP, DESERT, QUIT, LEAVE Of these seven verbs, so variously derived, though at first sight apparently synonymous, conversing does certainly better show the peculiar appropriation, than books, however learned; for whilst through them by study all due information may certainly be obtained, familiar talk tells us in half an hour -- That a man FORSAKES his mistress, ABANDONS all hope of regaining her lost esteem, RELINQUISHES his pretension in favour of another; GIVES UP a place of trust he held under the government, DESERTS his party, LEAVES his parents in affliction, and QUILTS the kingdom forever.'

'DESPONDENCY, HOPELESSNESS, DESPAIR form a sort of heart-rending climax rather than a parallel -- a climax, too, which time never fails of bringing to perfection. The last of these words implies a settled melancholy, I think, and is commonly succeeded by suicide.'

There is an occasional attempt at semantic links among the sets, Thus, BASE, LOW, SORDID (p. 34) are followed by BEAUTIFUL, HANDSOME, GRACEFUL, but this possible antithesis is the exception rather than the rule. Mrs Piozzi does not use Trusler's indexing or numbering, but alphabetizes on the basis of the first word in each set. None of the secondary words in her sets can be found if the first words (arbitrarily given) are not known. There is no reason for example, why ABANDON should precede FORSAKE in a set, but it does and so dictates the place of the set in the book. If inadequate on retrieval and relatedness

among sets, Piozzi compensates by indicating the collocational restrictions on words, as for example when she says: 'PAIR, COUPLE, BRACE - All mean two of one sort, yet cannot they be deemed true synonymes, while such arbitrary modes of using them prevail. A PAIR of eggs, or a COUPLE of coach-horses, would be ridiculous' (p. 103).

There have been many attempts at coping with synonymy since Trusler and Piozzi. The uncertainty of format in synonymies gradually gave way to a glossary style, and the features of the synonymies during the 19th century were increasingly carried over into standard dictionaries. William Perry in his all-embracing and influential Synonymous, Etymological and Pronouncing Dictionary of 1805 fused the traditions and dealt in resemblances between words. The utility of etymology in helping to discriminate word senses was emphasized in William Taylor's English Synonymes Discriminated (1813). Both works were however eclipsed by George Crabb's detailed English Synonyms Explained in Alphabetical Order, which appeared in 1816.

Crabb's is a word-finding list extending over 772 pages, using etymology and citation and discursive personal comment to justify his points. While not wishing to depreciate earlier efforts, he was not happy with English-language synonymies, which had 'fallen below other nations... whilst the French and Germans have had several considerable works on the subject, we have not a single writer who has treated it in a scientific manner adequate to its importance.' In the opinion of Rose Egan (1942 (1968:11a)) his main achievement was the ordering of sets from the most general to the least general term, but Emblen considers that Crabb's work was vitiated by excess: too many citations, too much personal opinion, too much faulty etymology. These are the

sins of the age, shared with Johnson and Webster, and perhaps the criticism of the Quarterly Review is more telling. It castigated Crabb as 'maddeningly circular'. Premeditation cross-refers to forethought which takes you to forecast which points to foresight which suggests a look at premeditation (Emblen, 265). From the methodological point of view, however, this is an advance on Piozzi, in which subsidiary words of any set can never be found directly. Crabb improved on Piozzi by giving set members as individual entries, thus:

to assist v. to help

When the user goes to help he finds the complete set laid out: help, assist, aid, succour, relieve, with the introductory comment: 'The idea of communicating to the advantage of another is common to all these terms. Help is the generic term, the rest specific.' This kind of logical procedure is still followed in dictionaries of synonyms, and Crabb's occasional grammatical observations are highly apposite (as in the set opinion, sentiment, notion, that 'we form opinions, we have sentiments, we get notions').

The Trusler-Piozzi-Crabb line is the basis of the modern synonymies which Egan studies in detail in the introduction to her 1942 Merriam-Webster compilation. Whatever has been published since, with the exception of Roget, appears on the whole to be refinement and rivalry within an established special tradition, a gift of the vocabulary to the glossary tradition.

In 1851 there appeared a modest work called A Selection of English Synonyms, by the second woman in the field, Elizabeth Jane Whately, the daughter of an Anglican bishop who has sometimes been erroneously

credited with producing the book. Whately makes explicit the observations of her predecessors when she says that what we must discriminate are in fact 'pseudo-synonyms'. She denies that there are true synonyms. Among her other points are the social distinction between words of Anglo-Saxon and Latin provenance, swine's flesh and freedom being 'less refined' than pork and liberty. She distinguished lexicography from philosophy and was concerned with words among words, not with words as representations of ideas, a point which Egan approves and which sets the synonymists apart from Bishop Wilkins. She omitted direct discussion of etymology because it led to a confusion of past and present, but appears to have made good use of it to establish certain distinctions. With regard to changes in a word's use over the years, she says: 'All these variations of meaning ... are valuable and curious, but though they may occasionally help us, they must not be allowed to influence our decisions with respect to the significations of words. Our question is, not what ought to be, or formerly was, but what now is.'

Egan insists (15a) that a work which came out the year after Whately -- The Thesaurus of Peter Mark Roget -- was not a synonymy in any valid sense. She expresses strong feelings on the point, noting that 'only when it is clear that the book purports to be a supplier of words -- technically a "word-finder" -- and nothing else, are we able to estimate correctly the heresy that has arisen out of its misunderstanding.' Egan also has strong views about heresy within the camp of the synonymists themselves, considering that in more recent times the Merriam-Webster Dictionary of Synonyms has kept the proper 'narrower' distinction of a few words with essentially the same meaning as against the dangerous 'broader' approach (found in their rivals, Funk and

Wagnall) which simply lists as a set words which in some way overlap one another. She blames James C Fernald (1838-1918) for this departure, commenting: 'Further synonymists there have been, some very good, some not quite so good, and some very bad; but they have all taken sides, either with those who support the traditional definition of synonym as one of two or more words having the same essential meaning or with those who favour its extension to one or two or more words which coincide in some part of their meaning. There has been no compromise: it might even be said that the break has scarcely been noticed. Nevertheless, it is apparent that, unless there be some clarification in definitions, especially of synonym and antonym, the prevailing popular misunderstanding will increase -- with what results no one can estimate.' (23a)

A relatively dispassionate look at Funk and Wagnall's Modern Guide to Synonyms and Related Words (1968) and Webster's New Dictionary of Synonyms (1968) suggests that although Miss Egan's subtle methodological distinction holds true, it is not surprising that an unsophisticated world has failed to notice it. Both works take short lists or sets of related words and discriminate them in short essays of the Trusler-Piozzi-Crabb type. They have nothing of the conventional-topic ordering in them, do not employ pictures or numbers, and require the alphabet as a primary aid to word-location. They have, in fact, as much in common as rival desk dictionaries, differing only in minor points.

2.4 Roget's Thesaurus

Egan rightly noted that Roget's concern was not the discrimination of synonyms. Rather, he was concerned with organizing the vocabulary of

English in terms of a logical classification of ideas. In the Introduction to his Thesaurus of English Words and Phrases (1852) there is a footnote which lists earlier works of logical ordering which Roget examined. They include an 1808 translation of a Sanskrit vocabulary organized on a Hindu world-view, a French work in 1797 which contained tabular schemes for categorizing the world, and, most significantly, a reference to Bishop Wilkins and his essay of 1668.

It is not surprising that Peter Mark Roget should have taken a sympathetic interest in the recondite works of John Wilkins: for 22 years Roget was secretary of the Royal Society, which had supported Wilkins. He admired the immense labour and ingenuity that had gone into Wilkins's Essay but regretted that it had proved too abstruse to be of practical use. Roget may have produced a work specifically for English, but his dream was ultimately the same as Wilkins': something universal and logical beyond the natural languages. Of this dream Roget says: 'The probable result of the construction of such a language would be its eventual adoption by every civilized nation; thus realizing that splendid aspiration of the philanthropists - the establishment of a Universal Language. However utopian such a project may appear to the present generation, and however abortive may have been the former endeavours of Bishop Wilkins and others to realize it, its accomplishment is surely not beset with greater difficulties than have impeded the progress to many other beneficial objects' (xxxv/vi, in Roget, ed. Dutch, 1962).

Later commentators might recommend the Thesaurus as invaluable for crossword-puzzlers, but Roget was aiming at translators, philologists, historians, lexicographers, writers, speakers, philosophers and

scientists, in the hope of contributing towards 'a golden age of union and harmony among the several nations and races of mankind' (xxxvi).

Trained in Edinburgh as a physician and practising for most of his life in London, Roget had a range of interests comparable to da Vinci, Bacon and Comenius. He was equally able in discussing ants, arsenic, library organization or words. 'From the beginning of his professional career,' says Emblen (255), 'Roget had toyed with a project that an objective observer might well have insisted lay quite outside the scope and powers of the Secretary of the Royal Society, the Fullerian Professor, the physiologist, the water expert, the chess master, the physician, or any of the other competencies attained by Peter Mark Roget. Neither literarily nor philologically was Roget equipped to create the work that has made his name as much a household word as that of Noah Webster.'

Emblen implies that a man of genius can expand where he will. The physician had been gathering lists of words since 1805, for his own benefit in writing, but it was not until his retiral at 70 that he seriously began the Thesaurus. Roget had however four cardinal interests which helped in just this kind of compilation:

- 1 He had contributed to many encyclopaedias, the foremost among which was the Encyclopaedia Britannica. In 1815 and 1817 he wrote a series of articles on physiology and other matters for the Supplement to the 4th, 5th and 6th editions. This opportunity had arisen through the recommendation of his Edinburgh professor of philosophy, Dugald Stewart, who had passed his name to the editor, Macvey Napier.
- 2 Throughout his career he was interested not only in the specificities of medicine, physiology and such other sciences but also in general

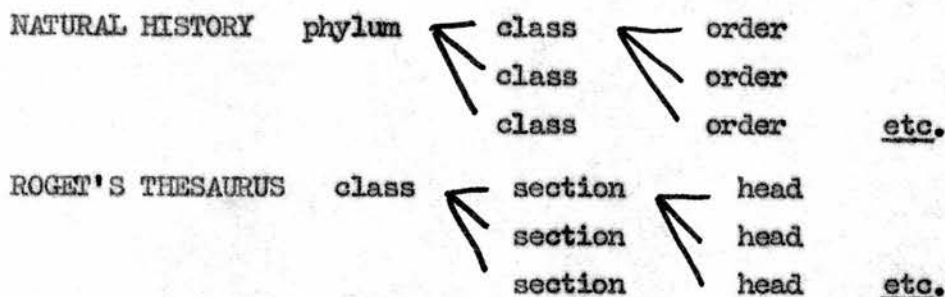
taxonomy. Hewas an organizer, a classifier in an age of classifiers, interested both in the taxonomy of natural science and the organization of books in libraries. He dealt with the practical problems of organization in the libraries of the Medical and Chirurgical Society and the Royal Society.

3 His interest in language was utilitarian, as a means towards clear expression. He was not interested in words or language per se.

Emblen for example (256) quotes his comments on a paper submitted for publication in a journal, which show his objection to turgid prose and careless spelling.

4 He was a perfectionist with a wider range of reading experience than the typical litterateur. Johnson, for example, suffered from inadequate knowledge of technical terms in context ('Ignorance, Madam, pure ignorance', he once conceded on the word pastern wrongly defined through his being unaware of specialized usage).

19th-century science was dominated by the botanical metaphor of the tree. This metaphor was used not only for the evolutionary growth of species ever upward (in Darwinian natural selection) but also for general classification. Roget acknowledges its influence when he says: 'The principle by which I have been guided in framing my verbal classification is the same as that which is employed in the various departments of Natural History. Thus the sectional divisions I have formed correspond to Natural Families in Botany and Zoology, and the filiation of words presents a network analogous to the natural filiation of plants and animals' (xxxv, footnote). Emblen directs our attention to this parallelism, and from his observations we can construct the following representation:

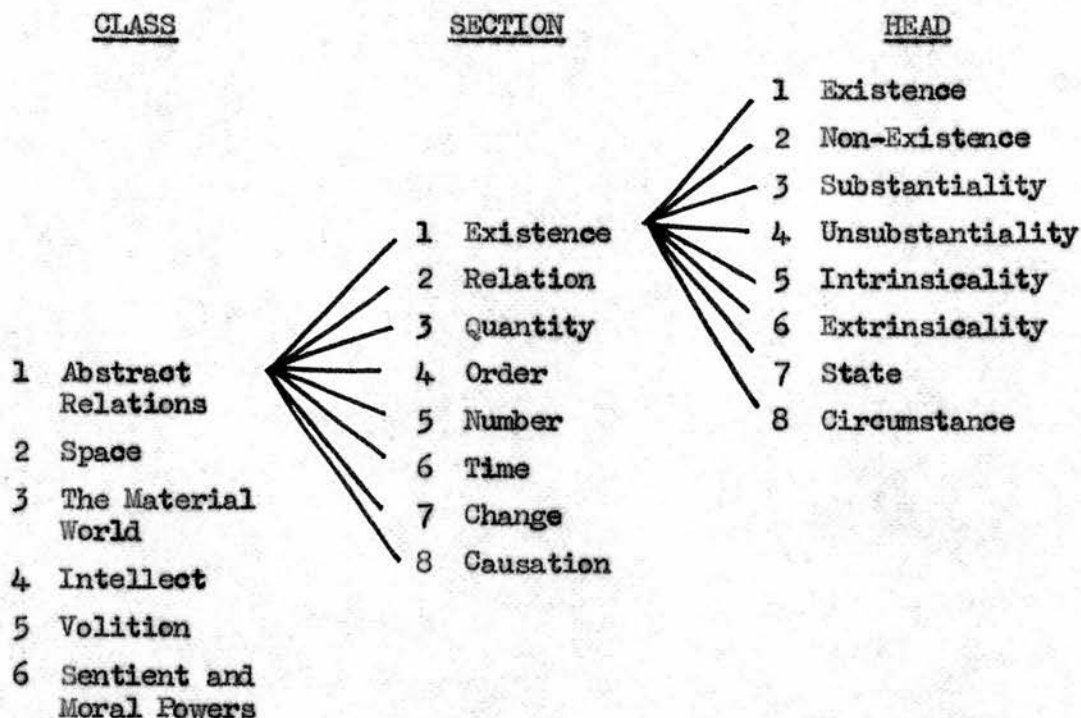


Roget combined the botanical taxonomy of the naturalists with the tradition of cosmic-ordering through the Aelfric-Withals-Comenius/Wilkins line. He wished to place a grid over reality, and label the appropriate nodes and branches. It is perhaps inevitable that a divine order of God and the angels should in the 19th century give way to biological phyla. It should be noted in passing, however, that even the powerful glossary tradition felt the impact of biological taxonomy. Murray, in the preface to the first volume of the NED (1888) similarly links his work with the taxonomists, but adds that no one should be deceived by the metaphor or by the taxonomy: 'For the convenience of classification, the naturalist may draw the line, which bounds a class or order, outside or inside of a particular form; but Nature has drawn it nowhere.'

Likewise, Roget drew his naturalist's lines across the language, but as his most recent British editor, Dutch, puts it: 'Language behaves like a continuum, coextensive with the thought it symbolizes, and does not lend itself to partitioning into self-contained categories' (1962:ix).

The 19th-century classifiers, Roget among them, were aware that they were erecting convenient fictions towards practical ends, but in the air of general optimism and progress these fictions have tended to

be taken as absolute, tempting us to forget the metaphor that underlies them, and to forget that any metaphor only takes us so far. Ultimately neither the natural world nor its creatures nor the words of a language are a branching tree. Given the convenience, however, Roget could propose a complex taxonomy, part of which is shown here:



Thus, existence is a head under the section existence in the class abstract relations. Man's relation to the world has complicated even since Comenius and Wilkins, let alone Aelfric, and hard 'facts' like beasts and fishes have vanished under abstract things like volition and extrinsicity.

Roget's abstract schema reflects a fact of English to which I have already referred several times: the special place which Latinate material occupies in the vocabulary of the language. Roget expresses his conceptual scheme almost entirely in Latinate terms, and a user might well be tempted to go to a hard-word dictionary like Cockeram's

to get a translation of extrinsicity, insalubrity or tergiversation. One can imagine Noah Webster castigating Samuel Johnson for putting such barbarous Latinisms in his dictionary. Roget however saw nothing cumbersome in any of his terms, insisting rather: 'I have accordingly adopted such principles of arrangement as appeared to me the simplest and most natural, and which would not require, either for their comprehension or application, any disciplined acumen, or depth of metaphysical or antiquarian lore' (xxvi). He saw Wilkins as abstruse and recondite, but did not feel the same charge could be laid at his own door.

Perhaps Roget had an idealized user in mind, a member of the educated Elect. Any user of the original Thesaurus was required to get to the marshalled ranks of words through the taxonomy: there was, at that time, no index, and Roget did not envisage any index. Egan comments: 'The modern consultant of the Thesaurus, accustomed to depending on the elaborate index (provided in 1879 by the compiler's son John L Roget), has little knowledge of the original plan of the book, though it has in no way been disturbed by revisers of the Roget family. But this plan is obviously hard to use and few consultants of the Thesaurus, if any, now avail themselves of it' (14a).

There have been many internal changes in the Thesaurus, even some special editions which re-worked the whole thing as an alphabetic synonymy. Deletions of many obsolescent terms, the addition of new material and the re-arrangement of contrasting lists have all served to up-date the original, but the taxonomy, as Egan observes, has been left severely alone. It is important to note this, because the taxonomy relates to a startling claim which Roget made for the wordbook,

and which his most recent British editor, Dutch, carries even further. It is, I think, a unique claim among wordbooks of English.

2.5 Words, Ideas and Thoughts

Roget says: 'The present work is intended to supply, with respect to the English language, a desideratum hitherto unsupplied in any language; namely, a collection of the words it contains and of the idiomatic combinations peculiar to it, arranged, not in alphabetical order as they are in a dictionary, but according to the ideas which they express' (xxiii).

Here we have thesaurus as converse of dictionary, words being reached not through their forms but through their meanings, through the 'idea' or 'concept' behind them. It follows from this claim that the taxonomic structure raised by Roget is not just a tree with word-labels at the nodes -- it is a structure modelling human thought, which happens to have Latinate labels attached for our convenience. This is an approach inherited directly from Bishop Wilkins, who was interested in expressing the ideas behind words in a scientific shorthand independent of words themselves. Roget wanted words to function as labels for something higher and more permanent. Dutch in his own preface defends the Roget position and the assumption that thesaurus and dictionary are complementary tools. In looking to a class or a section or a head, he maintains (viii), we are not looking at 'words' such as space and intrinsicity but at the idea beyond them. Words may change and have been changed in successive editions of Roget, but the ideas continue unchanged. Thus, the idea relative to clothes was once labelled investment but is now dressing. Dutch and Roget ask that we suspend

our knowledge that words are just words, and treat them as ideas when they label nodes in the tree. Then, once we get to the lists under each head, we can return to normal and take the offered words for what they are.

The organizational words at the nodes are not in consequence given their inherently interesting status as inclusive terms of academic English, but are labels used for want of some supra-linguistic shorthand for marking 'ideas'. Their rather remote Latinate nature may help in the necessary suspension of belief. This requirement relates crucially to a claim which Dutch makes explicitly for Roget, at two points in his preface:

1 'A thesaurus is operating on the same lines as a speaker or writer in the process of composition. It images in some measure the working of his (the user's) brain when, having his idea (corresponding to a thesaurus head), he mentally scans his stock of words (corresponding to the vocabulary of a thesaurus) for the right expression' (viii).

2 'It is the counterpart of the thesaurus we all carry in our memories in which mentally we track down a word. Surely, this characteristic is implied in those criticisms which impugn the merits of all thesauruses: that it is the lazy man's book; that it saves him the trouble of thinking. It would not do this unless it were patterned on our processes of thought and speech' (xiv).

In this claim the distinction between glossary and vocabulary is at its sharpest and most challenging (though vitiated by the user's dependence since 1879 on an index). Egan is probably right when she insists that few if any users consult the conceptual structure to get

a word neighbouring on the one they know. They use the index. To that extent (and it would be useful to get some hard facts of a statistical kind on just what Thesaurus-users do), the conceptual structure is irrelevant, and its truth-value as an analogue of the human brain is unimportant, just as the truth-value of Aelfric and Withals regarding God and the angels was irrelevant in retrieving Latin words.

If the Thesaurus is an analogue of the human mind, it is a sketchy one. Roget in his compilation achieved a masterpiece of specialized listing, but left the following to the user:

- 1 Providing the definitions. The user must know the meanings of the words when he finds them.
- 2 Making the discriminations. The user must know how the words differ and have the skill to use them appropriately.
- 3 Adding the derivatives and compounds. The user must know the principles of derivation in English and how far they can be pushed in coining new forms from established words.

These points did not trouble Roget. He had no intention of providing definitions, etymologies, discriminations, citations, recommendations or anything else current in orthodox lexicography. He refused to interpose himself between users and life as an arbiter of style, and had no wish to enter the labyrinth of nice distinctions between words esteemed synonymous. He assumed that his consultants, being competent users of English, could manipulate the material to good ends. His book was compiled as an aide-memoire for the Elect.

There is an important sense, however, in which Dutch is right, and the Thesaurus is the converse of a dictionary. Assuming that we cannot

escape the tyranny of words and that Roget's taxonomy is a tree and no more than that, it is still a sort of semantic ordering, an arrangement of words in associative catalogues which can be scanned for many purposes. The catalogues have been compiled by a 19th-century man and carry the inevitable mark of their time, but all in all they point to what happens in our minds even if they are not a strict analogue. Whatever we do in our minds, we do not have alphabetic lists of words. Perhaps we have clusters such as Roget indicates.

Modern information science with its enthusiasm for the thesaurus idea supports the possibility that a good arbitrary structure of the Roget type has its own justification, because it offers a usable system of retrieval. Emblen points to information science as a direct out-growth from Roget and the 19th-century classifiers(260). He notes that researchers and experimenters in such institutions as International Business Machines and Cornell University work on their own 'thesauri'. The classification of ideas into a skeleton outline of commonly understood terms is at the heart of the retrieval systems which are now operating or in an advanced stage of development. The new thesaurus is an authority list serving as a bridge between the searcher for any kind of information and the source of that information. Emblen quotes a letter from IBM which says: 'A thesaurus would be a structured list of words showing authorized terms, broad terms or narrow terms, "see also" references, and scope notes to explain the use of the term.' Stored in a computer, these lists can be updated regularly and referred to at a distance. One should note however that the absolute choice of the term 'thesaurus' has not yet been made by information scientists, and both in Emblen's own quotations and elsewhere, 'dictionary' is a competitor. The Random House Dictionary does however list a sense of

thesaurus (1966) under Computer Technology, as 'an index to information stored in a computer, consisting of a comprehensive list of subjects concerning which information may be retrieved by using the proper key terms.' Such a definition is not given in Webster's Third (1961). Emblen and the RHD may well be right, and 'thesaurus' will win against 'dictionary' in the computer world.

2.6 The 'Sacred' Thesaurus

The Thesaurus was compiled by a non-philologist, drew on biology for its taxonomy, suffered from an opaque Latinate conceptual system, was compiled for philosophical purposes with a world super-language in mind -- and was a commercial success from the start.

Crabb's work on synonyms was a great success, but Roget has dwarfed Crabb and consigned most of the synonymists to oblivion. So successful was the new compilation that it became a family industry over three generations of Rogets. Peter Mark Roget worked at it until he died at the age of 90 in 1869, having personally seen 25 editions through the press; his son John worked on it till his death in 1908, when it passed to the grandson, Samuel Romilly Roget, who carried it on till his death in 1952, when the British rights went to the publisher, Longman.

Emblen points to the 'near-worship' in which the great classifiers were held in the 19th century. A reviewer of four works on botany in 1852 (the year the Thesaurus came out) eulogized the taxonomists, uttering their names 'as an incantation, an "open sesame"' (259). Something of this aura of sanctity seems to have passed to the Thesaurus. Acclaim has become repetitive and increasingly uncritical, from the

early days when reviewers worried about how to use the volume, while approving of the labour that had gone into it. Modern reviewers tend to greet the latest revision as the up-dating of an old and trusted friend. Emblen is puzzled by this syndrome of success, largely because of the absence of serious examination of the thesaurus and the synonymy, and because of the almost total lack of interest in Roget the man, as opposed to Roget the institution. The eulogy of the classifier seems to have been diverted into a eulogy of the classification. The book dwarfs the compiler.

An early reviewer considered that anyone who could use a biblical concordance could use Roget and noted that 'the labour must have been immense, but the author's reward is sure. Roget will rank with Samuel Johnson as a literary instrument-maker of the first class' (the Westminster Review, April, 1853). Emblen is right however in being puzzled at the obscurity which has surrounded the man. Twenty million copies of the Thesaurus have been sold, but no biography appeared until 1970.

Divinization seems to be the fate of all successful wordbooks, along with an unwillingness to ask about the frail mortal lurking behind. Although talking about dictionaries, Twaddell makes a point that is relevant here: 'The lexicographer is especially protected by the very awe that his work inspires; he is not known as a collector and processor of word-lore; he is just a shadowy phantom behind the overpowering facade of "The Dictionary"' (1973:216). Roget too has been a phantom.

The Thesaurus however has been very real, as judged by the small war waged around it. Early in its history an American edition was

brought out, and over the years two separate traditions have developed, one the concern in Britain of Longman, the other largely in the care of the American publisher Crowell. As with its dictionaries, America has labelled the Thesaurus 'International', and in the course of time re-exported it to Britain as Collins's International Thesaurus, in direct competition with the up-dated British versions. It is generally agreed that the two versions and their spin-offs are quite distinct. Emblen charts their conflicting courses in some detail, and observes (282f): 'Even as American English has strayed far from the Queen's English, so the American and British thesauri have gone their separate ways. And we are left with another irony: now, nearly one hundred and twenty years after the Thesaurus first appeared, both American and British publishers claim to have preserved the essential Roget.'

Such claims are not unlike the claims of disputing sects, religious or political, each the repository of the pure original teachings of the messiah. Publishers and critics may not take wordbooks as far as patriarchs and caliphs take their doctrines, but the resemblance is strong. In the story of Roget's Thesaurus we now have two distinct revered traditions, unexamined because they are established institutions.

2.7 March's Thesaurus-Dictionary

The particularly successful features of various works in the vocabulary tradition have tended by and large to be incorporated into standard dictionaries, as we have seen with the pictorial aids developed by Comenius. We have also seen that the compilers of vocabularies have tended to want the security of the alphabet, so that synonymists such as Crabb and Egan have produced works which are, essentially, glossaries

of synonymous terms which can be found by reference to the first letter of the first word of a particular set of synonyms. Similarly, although Roget attempted to work only from a conceptual scheme, his son found it necessary to compromise with the expectations and abilities of readers, and introduce the index for which the Thesaurus is now famous. It has in fact proved difficult for the two traditions to stay apart, and it may be for this reason that historians of the subject have tended to disregard the secondary tradition of the vocabularists. It is important to bear in mind, however, that though the two styles may share their techniques, they differ fundamentally in the assumptions which underlie their compilation. The fact that lexicographers have preferred the alphabet as a basic tool should not serve to obscure this point. Versions of Roget may now exist in simple alphabetic format, but Roget's plan did not include alphabets and indexes.

The pressure of alphabetization can best be seen in the work of Francis March, an American philologist of considerable influence in the second half of the 19th century. March appreciated the distinctiveness of the two traditions in lexicography, having been himself closely involved as an American correspondent in gathering material for Murray's *NED*, a point which is acknowledged in the Historical Introduction to that Dictionary. It is not unfair to say that March and his fellow professors of English in the United States took the preparation of the *NED* more seriously than did their equivalent academics in Britain itself.

March was also, however, fascinated by the attempts of the synonymists and of Roget to find a solution to the problem of a word and the company it keeps, in the matter of synonymy and antonymy and also in definition. Clarence L Barnhart points out (1968:iv) that

more than 40 years before March offered his own solution, he had, in 1861, worked out a theoretical basis for a new kind of wordbook. March considered the study of synonymy 'one of the most valuable of intellectual disciplines, independently of its great importance as a guide to the right practical use of words.' For him it was an indispensable tool for achieving precision and accuracy of thought, an interest in logic and careful use of language which aligns him with Wilkins and Roget as well as with Egan. He felt, however, that Crabb's synonymy was flawed by bad etymology and that Whately's was too brief to be really useful. Roget in turn, while an exciting new achievement in lexicography, failed because it did not define and discriminate the words so conveniently juxtaposed.

Aided by his son, March attempted a blend of the glossary and the vocabulary traditions, and in 1902 published 'A Thesaurus Dictionary of the English language', a large single-volume work of subtle complexity. It went through five editions in its original form, and has since been revised and up-dated twice, in 1958 and 1968, particularly to accommodate technical usages.

Barnhart observes of the work (v): 'By combining the principles of the dictionary and the traditional thesaurus it was enabled to serve as a very practical reference tool for its users, whether as writer or reader. No doubt the experience of the younger March as assistant etymological editor on the staff of The Century Dictionary and Cyclopedia (1889-91) and as head of the etymological department of the Standard Dictionary (1893-95) contributed much to the practical side of the collaboration, but the theoretical basis for the book had been worked out by the elder March years before.' A study of the thesaurus-

dictionary, however, shows much more of the dictionary than any 'traditional' thesaurus. In effect, like many others, Barnhart means Roget when he talks about a traditional thesaurus, and by and large March rejects the idea that animated Roget and such predecessors as Wilkins and Bathe. He was not interested in structuring the cosmos. He was only interested in a reader getting to a word.

His method was simple but the result is bulky. He used the alphabet as his primary tool, so that like any dictionary the book is its own index. The user looks for a word. If for example we want information on the item arbiter we get a short entry as follows:

ar'-bi-ter. A judge. ADVICE, DECISION-MISJUDGMENT, JUDGE, MANAGER

Here March provides a short definition by synonym; elsewhere he might give a brief phrase. The meat of the matter, however, comes in the capitalized single and double words which tell users to go to such other points in the book if they want more information. Thus, going to ADVICE yields:

ad-vice'. Counsel. ADVICE, ENLIGHTENMENT-SECRECY, MOTIVE-DEHORTATION, TIDINGS-MYSTERY

This similar set of definition plus further suggestions is immediately followed by the capitalized heading ADVICE, followed by a sublist inserted in the main alphabetic list, containing its own sub-alphabet, a part of which runs:

Adhortation. Advice
Admonition. Friendly counseling or warning. See WARNING.
Advocacy. The act of pleading for, defending of.
Charge. An instruction; advice to.
Conference. The act of advising together.
Consultation. The act of consulting, of deliberating.

The basic list is followed by several others which add nouns of instrument such as chart, guide and manual, nouns of agent such as

adviser, arbiter, archiater, council and councilor, then figurative expressions, phrases, verbs of advising, adjectives, interjections, Latin phrases, and then we are back into the main alphabetic list once more.

It will be recalled that under arbiter we were also offered a double reference DECISION-MISJUDGMENT. If we referred ourselves there, we would find a simple definition of decision with its own further suggestions, then the headed sublists for DECISION-MISJUDGMENT, the synonyms given in a column on the left and the antonyms running opposite on the right, the whole appearing within the normal alphabetic flow, thus:

DECISION-MISJUDGMENT

Adjudication.	The act of rendering judgment.	Fixed idea.	An idea firmly established as by prejudice.
Appreciation.	True and adequate judgment or estimation.	Fool's paradise.	Misjudged happiness.
Arbitrament.	Act of judging by arbitration.	Foregone conclusion.	A conclusion made up beforehand.
Arbitration.	Hearing and judging a question by a party mutually agreed upon by the interested ones.	Hasty conclusion.	A conclusion arrived at without due deliberation.

The two columns do not relate directly to each other. The user simply runs his eye down each side for what it contains. We are dealing here with antonymous concepts, not individual pairs of antonyms.

Several points emerge from this brief description of how March's massive thesaurus-dictionary works in practice:

- 1 It is cumbersome and repetitive; the same expressions re-appear again and again in different places in order to save the reader from chasing through the book, and yet the multiplicity of capitalized

suggestions still send the user in just such an endless voyage of discovery. The user in fact decides when he has had enough. The system has no natural cut-off point in the provision of information. Because of the amount of repetition and listing, however, there is little room for detailed defining and discriminating. The explanations of the innumerable words are very brief and the number of capitalized suggestions very large.

2 As with most works which seek to provide generic terms, singly or in pairs, it has to rely on obscure Latinisms to fill gaps that have been left in the ordinary language. In such cases, while many generic terms and pairs are excellent, others are somewhat unusual:

CHOICE-NEUTRALITY
MULTIPLICITY-PAUCITY
SUPREMACY-SUBORDINACY
RENOVATION-RELAPSE
CRY-ULULATION
COLOR-ACHROMATISM
ODOR-INODOROUSNESS

3 March has, like many lexicographers, an enthusiasm for completeness, and the contents of the wordbook are swelled with words that would readily have qualified for inclusion in a 17th-century hard-word dictionary: nidor, oinomania, loup-garou, machicolation, graphoideophrenia.

Of the work, however, Norman Cousins (1968:vi) claims: 'Seldom have I met a March's Thesaurus-Dictionary owner who didn't express a devotion to the book verging on a literary mission to proclaim its virtues. When March's went out of print during World War Two, its users constituted something of a proud and possessive band of the lexicographically-privileged. For they regarded March's Thesaurus-

Dictionary as one of the three or four handiest and most valuable reference books about words in the English language. Speaking personally, I can attest that there is no wordbook in my own library which has served me better or which I prize more highly than March's.'

Unlike Roget and Webster, however, March has not succeeded in becoming one of those lexicographical names which is almost a household word.

2.8 Conclusion

For the greater part of their history, wordbooks in the vocabulary tradition have depended on specific representations of the cosmos, as understood variously by mediaeval clerics, Renaissance schoolmasters, scholar-scientists of the Enlightenment and ultimately by a remarkable Victorian polymath. Inherent in the tradition is a philosophical approach to concepts, best expressed perhaps in Emblen's comment on Peter Mark Roget: 'We must thus see the Thesaurus as a philosophical effort, not for the sake of a philosophical exercise, but in order to recognise in it the ultimate signature of a man whose very intellectual and emotional being demanded an ordered and a charted (or at least a chartable) universe' (261f). In this, Roget was the 19th-century culmination of the tradition of Aslfric, Withals, Bathe and Wilkins.

Beginning as a simple pedagogic device, the vocabulary has always had the tendency to appeal to philosophers like Wilkins as well as schoolmasters like Greenwood. Whatever lists of conventional topics were drawn up, however, we see in them not just a subjective grid placed on reality, but an indication of how a particular culture at a particular

time saw its world. This kind of involvement of a retrieval device in the general human predicament is quite different from the glossary tradition, which escapes philosophy by taking refuge in the apparent security of the arbitrary alphabet. Out of the glossary tradition there could hardly emerge an attempt such as Wilkins to find a universal system of language and writing that would transcend the limits both of languages and alphabets.

Close to this general tradition of cosmic-ordering there flourished other practical interests. Two such interests were the concern of Comenius for pictorial presentation, and the interest of the Trusler-Piozzi-Crabb tradition in discriminating words that appear initially to be synonymous. Both subtraditions have been fully exploited since they were introduced into English lexicography, but have tended to pass rather to the glossary tradition than to remain in the conceptual schemes of such compilers as Roget.

Like the glossary tradition, the vocabulary tradition has had its internal disputes, such as the subtleties of argument over what is a synonym, and the war between the successors to Roget. The tradition has not given birth to the sheer quantity of works that are found in the dictionary world, but the remarkable sales successes of the vocabularies should not be ignored. They are an indication that innovation also pays.

Both the glossary and the vocabulary tradition began in the mediaeval battle to learn Latin, the medium of culture and religion. Although each has expanded to cover the general words of the English language, they began as bilingual exercises and as vehicles for the

acquisition of the difficult, the abstruse and the odd. To some extent the majority of wordbooks today, from the NED to Roget, remain élitist, available only to the motivated few. It is probably not surprising, therefore, that at the end of the 19th century a third tradition arose that was concerned with wider matters, with vocabulary at work in the world, and particularly in various aspects of general education. This is the vocabulary-control movement.

3 The Vocabulary-Control Movement in the English Language

3.1 Introduction

The glossary tradition today is represented by a wide range of products: etymological, universal, unabridged, abridged, standard, concise, pocket, desk, collegiate, idiomatic, encyclop(a)edic, illustrated, dialectal, specialist, technical, unilingual, bilingual and so on through the spectrum. It has even acquired a UNESCO-sponsored compilers' guide in Ladislav Zgusta's Manual of Lexicography (Mouton, 1971), which takes the view that the glossary tradition is lexicography.

The vocabulary tradition has not blossomed comparably, and has passed many of its innovations on to the glossaries. Its finest current realizations are probably Roget and Duden. These do not at first sight appear to have a lot in common, until contrasted with dictionaries proper.

Both traditions have been affected (or can in future be affected) by the consequences of the more recent vocabulary-control movement which has had a short, almost meteoric career. It had its zenith in the 1930s and now appears to be falling away. In this short period it has however left its mark in both positive and negative contributions to word study in English. Its main historians are Bongers (1947) and Fries and Traver (1950). From these, some other sources and the contributions of the originators of the movement I shall try to show in this chapter how practical problems have propelled investigators to the brink of a unified theory of words.

3.2 The Special Word List

The concept of vocabulary control has been dominated since its inception by simple arithmetic and by a concern for words as chunks of lettering. More significantly than most kinds of language analysis, it depends on the printed word.

Mary Abercrombie has pointed out (personal communication, 1976) that simple word-counts in English date from the beginning of the 19th century, although such commentators as Fries and Traver tend to see the movement beginning just as that century came to an end. Thus, in the Phonotypic Journal of October 1843 we find evidence that serious if informal work was being done by both stenographers and educationists on frequency counts throughout the first half of the 19th century.

Isaac Pitman, in that issue of his journal, presented stenography with a list of words 'showing how often each occurs in 10,000 words, taken from 20 books, 500 from each... to serve as a guide in selecting grammalogues for the (shorthand) system.' Grammalogues are words which for stenographic convenience can be expressed as a single symbol (such as and, the, of etc.). The following list, taken from Pitman's longer alphabetic list, shows the items under S with their frequency ratings:

Saw	2	so	25
say	13	spirit	} 13
sea	} 24	spiritual	
see		stand	6
shall	34	strength	2
short	2	sure	2

Pitman does not state what texts were chosen for this survey, but even the brief extract above can tell us a lot about his approach.

We can see that no definitions were offered for discrimination, and so saw can be either a tool or a past participle of see, or both. Similarly, the homophones sea and see are conflated, presumably because in speech they are identical and in context no stenographer could confuse them. This means that Pitman was working simultaneously with three kinds of word: a graphic form or type, its tokens or the occasions on which the type is repeated, and the parallel in sound. We could call these graphological words, statistical words and phonological words respectively. In conflating spirit and spiritual, however, we see something else at work, namely the running together of two morphologically distinct but related words. The Pitman exercise is revealing in its very informality, its lack of theory, and yet at the same time its practical worth. In addition the list contains some interesting items such as betwixt and ought. These suggest that, like the conventional topics in the vocabularies, frequency lists tend to carry clues indicating the period in which they were compiled.

The practical worth of the list was interestingly highlighted in the November edition of the same journal, where a letter from a James Biden points out that a similar kind of list was made over 20 years earlier by John Freeman in London. Biden observed 'the two lists so strikingly to coincide that I make free to post you a copy of all the words that occurred more than 19 times in 20,000, as a strong and gratifying confirmation of the correctness of your own laborious calculation.' He worked from a second edition of Freeman in 1820, and adds that Freeman's work was undertaken as an aid to teaching adults to read.

Credit is normally given to the work in Germany of F.W. Kaeding (Haufigkeitwörterbuch der deutschen Sprache) for the development of frequency counting with a rigour and on a scale which would give it a serious objective value. Also with stenographers in mind, Kaeding organized the counting of eleven million words from a variety of texts and reduced these to their word types in alphabetic lists stating the frequency of occurrence of each. The first comparable modern counterpart to this in the English language occurred in the United States (1923), where George Dewey's The Relativ (sic) Frequency of English Speech Sounds attempted to establish for stenographers the commonest syllables for English. This work was also based on written presentation and is minuscule when compared with the German work, but larger than Freeman's and Pitman's. He took 100,000 statistical words from a scatter of novels, newspapers, speeches and the like.

The idea slowly became established that the techniques of such counters had a valuable advantage over the more subjective collections of the traditional lexicographers. The idea of counting tokens in a given corpus of texts, to establish undeniable patterns of frequency of occurrence proved very attractive, especially in the United States.

Educational word counts took several forms. Workers set out to organize objective lists that would help pupils in their reading, spelling, pronunciation and general use of language, hopefully replacing the old-fashioned little spellers of those who had followed Nathaniel Bailey and Noah Webster. The Rev J Knowles in Britain in 1904 produced his London Point System of Reading for the Blind, which took 100,000 words from the Bible and other sources and abstracted the 353 commonest words. The first 72 of these were grammatical, such as pronouns and

prepositions, the 73rd being the noun man and the 353rd being the noun child. Knowles claimed that his list provided three-quarters of all the words occurring in the texts assessed. In 1914 in the United States Cook and O'Shea undertook a survey related to the spelling needs of American children, in which the family correspondence of thirteen adults was chosen as a corpus. It was found that nine words accounted for over a quarter of the total, and that 42 accounted for over one half. The observers considered that 763 words took in 90-91% of the total correspondence.

Another line of interest emerges in Six Thousand Common English Words, by R C Eldridge in 1911. These words were drawn from a small count of 43,000 words in American newspapers. They were garnered as a contribution towards a 'universal eclectic or polyglot vocabulary', and in this we meet again the old search for the special universal language which will surmount the difficulties of natural speech and writing.

Stenography, reading, spelling and the search for a universal language led inevitably to larger questions in language teaching, both for native users and for foreign learners of English. The initial successes made for optimism.

3.3 Thorndike: the Objective Approach

In 1921, Edward L Thorndike brought out, in the United States, The Teacher's Word Book, a list of 10,000 words that an American child could expect to meet in his general reading. Thorndike, to get his words, used the basic frequency approach but added a range rating,

so that the questions 'How often?' and 'In how many texts?' could both be answered.

Thorndike's list was derived from 41 different sources providing four million 'running words', that is, graphological tokens. Three million came from the Bible and the English Classics, half a million from letters, 300,000 from elementary school-readers, 90,000 from newspapers and 50,000 from general reading. The list was widely acclaimed as a breakthrough in vocabulary control and inspired numerous imitators and developers. It has chiefly been used as an objective measure of the appropriateness of vocabularies in schoolbooks, and as a basis for the construction of achievement tests in vocabulary, reading and spelling. Although not intended for such use, it has also been taken as a basis for word lists for foreigners learning English.

This 1921 list embodies the basic assumptions of the objective counters. It relates entirely to written language, and behind the attractively objective arithmetic is a subjective or culture-based selection of texts from which the counting can proceed. The list rests on the tacit 19th and early 20th-century assumption that the Bible, certain 'Classics' and a random selection from appropriate people's letters and newspapers constitute a proper sample of English-at-large, or at least the best kind of English. Such a corpus either would or should serve as a guide to what readers may be required to know at certain ages and for certain examination purposes. In Thorndike therefore we have behind the objectivity an implicit interventionism, an unconscious attempt to perpetuate what is taken to be a norm. The Bible's presence in the count highlights this, much as the

presence of angels distinguishes Aelfric from Duden. It is interesting to speculate whether a count made, say, in 1980 would place the same emphasis on the Bible.

In 1931, Thorndike published an extension of the list to 20,000 words, by incorporating material from over 20 other sources, including about five million running words from other counts. This move typifies later developments in the objectivist tradition: the enlargement of lists by the integration of smaller lists and the standardization of arithmetical anomalies to this end. It was assumed that such incorporation would cast the net wider and mean greater accuracy. In the second list, words were rated from 1 to 20 according to frequency and range. In looking at this kind of development, Widdowson (1968:133f) notes that such expansion tends to cancel out important differences among words. Something counted as the same word in the Bible, one or two Classics and a scientific text may have nuances of meaning and use which cannot be exhibited by so blunt a method: 'This neutralizing of individual differences comes as a result of concentrating on comprehensiveness at the expense of discrimination.' Just how blunt the instrument was however emerges in the fact that for arithmetical convenience Thorndike allowed his counters to treat traditional homographs as the same. Thus, bear = animal, and bear = carry are lumped together. A frequency-range rating for such items is meaningless.

The willingness of Thorndike and his collaborators to conflate lists is additionally disturbing because of the need to seek common criteria for different counts. Some counts for example took is, am, are, was, were and be as six different words, while others treated them

as variants of the one word be. Conflation demanded complicated re-casting to bring lists into alignment, and implied that counters should re-think the whole basis of counting. If even one major fault emerged in later reviews of earlier counts (as for example the homograph situation) then logically it demanded a re-count, a return to first principles, having learned from one's mistakes. This however never happened. Counters were content to tinker with counts already made, probably because of the sheer difficulty of making counts in the days before computers could absorb the tedium. The assumption appears to have been that any count was better than no count at all.

Thorndike did not assume that his lists were absolute, any more than lexicographers have assumed that they carried the Mosaic Tablets. There is evidence however (in the way the lists were used) to suggest that educators tended to treat Thorndike's List as absolute, and the likelihood that people will always do this kind of thing only emphasizes the need to make sure that first principles are as clearcut as possible. The difficulty in the objectivist tradition lies in discerning any clear first principles at all, apart from the view that words are chunks of lettering with white space on each side.

Of his first attempt, Thorndike observes: 'The gist of it is that the present counts are adequate to determine the first 1,000 words with a small probable error, and the next 4,000 well enough for many educational purposes, and the last 5,000 to an extent that is useful, though far from accurate' (1921B:348). By 1931 however he had doubts about the ultimate accuracy of his lists, feeling that the rarer words were still not properly placed, some words having lower credits 'than they should have' (quoted in Fries and Traver (FT):25). This comment

raises a query about the theoretical/philosophical assumptions behind the count. What does should imply? Does it mean that the lists are not really worth much in their objectivity, because subjective intuition suggests that they do not tell us anything worthwhile? Or does the self-criticism mean that Thorndike measured himself against some Platonic Ideal List, expressing frequencies and ranges known only to God?

The many other counters added little to Thorndike. The same words re-emerge in the first 500 of each list, similar to the far less sophisticated efforts of Freeman, Pitman, Knowles and Cook and O'Shea. More and more the statisticians questioned their findings about the rarer or 'special' words and urged the accuracy of that first 500. Since the first 500 consists very largely of the grammatical words of the language this is little more than an affirmation that English has a grammatical structure which uses grammatical words. The counters Faucett and Maki (1932) moved away from a pure objectivist stance by urging that personal judgment was needed for 'wide-range words', and so the objective lists can be adversely criticized from both ends, the commonest and the rarest.

From 1934 onwards Thorndike turned to the problems of the homograph and the polysemic word, that is, towards some resolution of the problem of one chunk having different senses. With Irving Lorge he embarked on a semantic count which was related to the senses of words enumerated in the NED. Here, for the first time, the objectivists turned to a traditional wordbook for help. If successful, such a count could give an overall frequency for a chunk of lettering such as stenographers and spellers might need, and also sense frequencies for

specialists. No inquiry was made into the absoluteness or otherwise of the NED division of word-meanings into senses, and we can therefore take it that Thorndike and Lorge were interested in a useful construct rather than a truly objective measure of English lexis. If this is so, then the undertaking is the finest hour of the movement. It is certainly indicated (in West, 1953:xii) that the counters in this undertaking were linguistically sophisticated (probably in much the same way as the volunteer readers for the NED) and the results are impressive. What we have are precise statements about printed words taken from a culture-related selection of texts charted in accordance with the senses numbered in a historical dictionary. Such a count can be taken as an indicator of how words appear in written English, as a supplement to the judicious decisions of, say, a textbook-writer, but it cannot serve as an ultimate authority.

In 1944 appeared the third and last extension: Thorndike and Lorge's The Teacher's Word Book of 30,000 Words, which had added in new material from Lorge's magazine count and Thorndike's juvenile count, together with material from the semantic count. The earlier criticisms still hold good, along with one not previously raised: the massive presence of personal and place names. On one page of the 1944 list, for example (randomly selected), the following appear:

Massachusetts	Maximilian	McCormick
Matilda	May Day	McDonald
Matt	Mayo	McKinley
Matthew	Maypole	Meade
Maud(e)	McCarthy	Mecca
Max	McClellan	

Many questions could be asked about these, each meticulously provided with its frequency and range. They are hardly words of

pedagogic interest outside geography and history. In their massive presence they point to the basic misassumption of starting to count without asking what should be counted. They also point to the bias of the texts chosen, and ambiguity in the listing. Is Mayo the place or a person or both? Can McClellan be given a frequency of any kind, especially at the expense of all the other Mcs and Macs in Scottish and Irish telephone directories or even of the spellings McLellan and McClelland? Can it be seriously claimed that Massachusetts and Mecca are more worthy of frequency rating than Majorca and Melbourne, which are off-listed?

The literary bias of the objective counters was inevitable, but led, as Bright and McGregor point out (1970:17f), to peculiar results of some importance to the teachers they were serving: 'The tape recorder had not then been invented. One result of this (reliance on the written word) is that some everyday household words have an oddly low frequency. Sticky is as rare as strew and doorknob as rare as dulcet. Moreover the list is now out of date. The word transistor does not appear at all, and the frequency of pop takes no account of this graphic symbol's recent semantic extension. The list was designed for use in American schools and consequently has a slight American bias but even so trousers is much more frequent than pants... Damsel and doth are as frequent as error and encounter.'

Of mammoth proportions, the statistical approach appears to have been largely misapplied, its contribution being negative rather than positive. If a count were organized on the basis of a theory of words which omitted the grammatical material, categorized what should be taken as lexical and worth counting, and then proceeded for a carefully

selected corpus (say, of medical or some other specific kind of language), a statistically interesting and practical list might emerge. The blanket approach of the objectivists however serves only to emphasize the adage that what you get out is only as good as what you put in.

3.4 Palmer and West: the Subjective Approach

At the same time as the American interest in frequency counts was developing, British teachers of language were experimenting with lists which could serve as bases for simplified readers, textbooks and general courses.

The foremost figure in the early stages of the British tradition was Harold E Palmer, working first in Belgium, then London, and then, for the most productive period, in Japan. In his early work as a language teacher, Palmer was impressed by the elementary word lists used in the Berlitz method of modern language teaching. In looking at the Eldridge list of 1911, he suspected the worthwhileness of frequency counts and sought instead some view of words which would make listing systematic while still subjectively based. To this end, he developed a system of three 'lexicological units':

- 1 the monolog, something neither more nor less than a word
- 2 the miology, something less than a word
- 3 the pliolog, something more than a word

This typology, as Widdowson points out (131), sidesteps the problem of what a word is, and mixes criteria, the monolog being a kind of element or substance, the miolog being something dependent on the

monolog (an affix perhaps), and the pliolog being something structurally complex (like an idiomatic or set phrase). The typology glosses over many problems, but it is a step towards deciding what is worth selecting or even counting. In practical terms, Palmer added to this typology a dichotomy of headword and subword. In any list, the headword is the main entry, and the subwords are the various pliologs associated with it (derivatives, compounds, and idioms). An example of the method is the following, from Bongers' K list, which uses Palmerian principles:

absent adj
 absence n
 absent-minded adj
 accept v
 acceptance n
 acceptable adj
 unacceptable adj
 accident n
 accidental adj
 accidentally adv

Bongers claims that Palmer worked essentially from a 'vocabulary sense' developed over years of teaching, a sense which 'enabled him to determine with little hesitation the relative utility of any word ... a typical example of the Subjective or Empirical Method of Vocabulary Selection' (74). When Palmer went to Japan in 1922 as linguistic adviser to the Japanese government, he developed his methods through the medium of the Institute of Research in English Teaching (IRET), under whose auspices he prepared two interim reports. In 1934 Bongers travelled from the Dutch East Indies to Tokyo to meet him, initiating a long period of co-operation. 'Palmer believes (and so do we) that subjective judgment and the empirical evidence of teachers is of invaluable assistance in compiling word lists that will serve more adequately than lists based upon objective findings alone' (45).

Palmer's approach, as laid out in the first interim IRET report (1930), received the blessing of I.A. Richards, who wrote: 'Determinations of the relative frequencies of words in selected bodies of literature give us at best only raw material. We need to know the assumptions on which they are made' (204). In the report, Palmer offered certain assumptions about words which amount to a theory for their selection. The whole set of principles is not given here, but an abstraction of ten highlights the distinctiveness of Palmer's approach:

- 1 No absolute statements should be made about certain words being the important words of English, but only that they seem to be more important than others.
- 2 Inflected forms are simply variants of a given word.
- 3 Head words should not include miologs or pliologs.
- 4 Variant forms, abbreviations and the like should not count as separate items.
- 5 Words of wider range should be taken to include related words of narrower range.
- 6 Stylistic words should be excluded.
- 7 Homonyms are separate words, and senses of a given word should only be differentiated when they are so extreme as to be 'quasi-homonyms'.
- 8 Selected vocabulary should be drawn up in a general list and specific lists.
- 9 Each list should be given an arbitrary numerical limit (its radius). The words within that limit would constitute a zone, say, a zone of the first hundred units, or first thousand, and so on. The supplementary lists would relate to specialized zones.
- 10 Each unit in the general list is a headword together with any subwords.

With these axioms in mind, the compiler of lists can proceed to use his 'vocabulary sense', or intuition.

Objectivists criticized Palmer just as he criticized them, arguing that there was no external control on his choice. With misgivings, he compromised far enough to incorporate the first 500 words of Thorndike into his future work, but in the second interim IRET report (1931) he defended reasoned as against mechanical selection by calling attention to what he called 'constellations' of words. 'The lesson in which the word eat occurs for the first time contains also the word drink.' In such a relationship frequency of occurrence is the product, the outcome, not the promoting factor. The constellation will include hungry and thirsty and some 40 other words linked by 'thought association'. He denied that many of these associated words would appear in a frequency count at anywhere near the right rating. He did not however develop the idea of constellations in their own right with their internal relationships marked, preferring instead his zones, expressed as alphabetic lists.

In the course of a world tour in 1931, when he met many of the foremost figures in the vocabulary-control movement, Palmer met Michael West, who had worked on word lists in Bengal and for his New Method Readers, written for foreigners learning English. The vocabularies of these readers, subjectively achieved, had served as the starting point for a refining procedure, an eclectic sampling and comparison of both subjective and objective lists until West achieved a special 'defining vocabulary' of 1,490 items. This metalanguage for definition is of special lexicographic interest because it was put to use in a wordbook: West's New Method Dictionary (Longman, 1935) and in the later derived An International Reader's Dictionary (1965), which defines over 24,000 items within this vocabulary. The preface of the second dictionary states that the 1,490 words are 'held to be the commonest words in

English or the words first learnt by foreigners.' If the list is derived from an influential series of simplified readers, on which other simplifications are founded, then the statement is a self-fulfilling prophecy: the day will arrive when it is largely true because it has been made largely true.

It is with West that the word lists join the mainstream of the glossary tradition, and colleagues of Palmer followed him with The Advanced Learner's Dictionary of Current English (Oxford, 1948 onward) by A.S. Hornby, E.V. Gatenby and H. Wakefield. It began its life as an IRET work, first published under another name, Idiomatic and Syntactic English Dictionary, in Tokyo in 1942. No claim is made for the ALDCE with regard to a finite defining vocabulary, but the defining words used could be added up to make a list like West's. They have been intuitively and pragmatically gathered for the purpose of defining, and are intentionally 'as simple as possible' (ALDCE, vi).

The meeting of Palmer and West was followed in 1934 by a specially-convened conference in New York, under the auspices of the Carnegie Corporation. West was instrumental in bringing it about, to allow the various objective and subjective compilers to meet and share ideas. The conference was the zenith of the movement, but should not be discussed before mention is made of a controversy in which West was deeply involved, and one which reflects rather sadly on the whole business of vocabulary control. In it, the peculiar fascination of a universal language re-appears.

3.5 Ogden: the Logical Approach

C.K. Ogden developed 'Basic English' in the 1920s, and whether discussed in a lexicological context or within the framework of artificial languages like Esperanto, his work remains as challenging today as it was in the 20s and 30s. There are many reasons for including Basic in this study: firstly, it is linked with an important work on philosophy and semantics (The Meaning of Meaning, 1923, of which Ogden was co-author with I.A. Richards); secondly, its universalism links it with Wilkins and Roget and also with Eldridge, one of the early word-counters; and, finally, there are the claims of the creator of Basic as to how it can be used.

Basic can be regarded as an exercise in applied linguistics and language planning. It is the logical extraction from a natural language of material assumed to be useful for international communication. The term 'Basic' is an acronym with the values B for British, A for American, S for Scientific, I for International and C for Commercial, and to the acronym we can add the three claims which Ogden made on its behalf:

- 1 It is a world auxiliary language in its own right.
- 2 It is also a way of leading non-English users to an understanding of normal English.
- 3 It is a means of improving the critical awareness of those already using normal English, reminding them of the virtues of simplicity and precision.

To achieve this three-in-one language goal, Ogden isolated from 'Standard English' a simple syntax and a logically restricted vocabulary. The basic sentence or clause type had a fixed analytic order: I will put the record on the machine now. Six affixes could be used: -s for plurals, un- for negating adjectives, -ed and -ing with their traditional

function in making participles, -ly for adverbs and -er as a highly productive agent suffix. In addition, permissible word compounding included:

- 1 phrasal verbs (that is verb plus adverbial particle structures like get up, put on)
- 2 noun + noun structures (like teapot, farmhouse)
- 3 adjective + noun structures (like madman, blackbird).

The minimal syntax was buttressed with 850 words, divided into 400 general words, 200 picturable words, 150 quality words, 82 grammatical words and a special set of verb operators of the type give and set, which could serve as the basis for compounding and also for such combinations as, say, give him a push instead of push him. Ogden allowed for the possibility of metaphoric extensions of the basic meanings of the 850 words, and accepted also that they could be supplemented by four further categories of words:

- 1 numbers
- 2 names of local animals and plants
- 3 proper names
- 4 scientific-cum-international words as the need arises.

The audacity of this artifact, separated out from the living language, obtained a controversial response. Perhaps its somewhat evangelistic aim of being all things to all men made it an easy target for attack. Ogden sought to resolve a multitude of problems at a stroke, combining a spirit of progress with a kind of interventionism, on the assumption that people's use of Standard English was not all that it might be. It should be borne in mind, however, that the logico-semantic approach is only Palmer and West's subjective method pushed to its furthest limits, and the simplicity of the end-product is impressive.

Essentially criticism of Basic has been three-fold:

- 1 That you cannot combine an independent world auxiliary derived from English with a medium serving as a way-in to Standard English.
- 2 That Basic's dependence on minimal operators and combinations leads to circumlocutions unacceptable in Standard, difficult for a native user of English to master, distorting for foreign learners and ultimately leading to a debasement of the proper language.
- 3 That assumptions about a minimal vocabulary obscure the polysemic and special idiomatic significance in Standard English of its commoner words. That is, that the 850 words are not as simple as they might appear at first sight to be.

The third criticism of Basic is probably the most telling. Compounds and phrasal verbs in Basic often have overall meanings which are not entirely deducible from their parts, as in put up with meaning tolerate. Such meanings are idiomatic extensions, as it were, of the Basic meanings of these words, and can be taught by various devices such as indicating metaphoric extensions in special situations. It remains a problem, however, to decide where Basic meanings for a word stop and non-Basic meanings begin. Fries and Traver, for example, (80f) have shown that the 850 words of Basic have no less than 18,416 senses as listed in the NED. They suggest that Basic did not adequately come to terms with polysemy, the radiation of meaning in different contexts.

The simplicity of these Basic words is more apparent than real. Ogden is dealing largely with an Anglo-Saxon stratum, a kind of nativistic return to the simplicities of an age earlier than the Renaissance. His choice of structures (phrasal verb and nominal com-

pounds) is Germanic, and his active-verb sentences turn away from the passivization common with Latinized English. Like the 17th-century lexicographer Cawdrey he would transform Latin verbs like impose into phrasal verbs like 'lay on'. It is doubtful, however, whether a phrasal verb is really simpler in its syntactic behaviour or semantic intricacy than a Latinate verb, although at first glance on paper it looks simpler to the native user, who has been familiar with such expressions since early childhood.

Michael West seriously opposed the propagation of Basic, fearing alike its success or failure. Its success, he thought, would imperil more acceptable forms of simplified English and favour a gross pidginization of the language; its failure would cast doubt on the systems ~~as~~ he himself advocated. West's A Critical Examination of Basic English (1934) and Ogden's Counter-Offensive (1935) show how emotive the struggle became, West seeing Ogden doing mankind 'an incalculably grave disservice', while Ogden accused West of 'gross errors' and 'ludicrous' criticism in his assessment of Basic. Ogden turned down an invitation to the West-inspired conference in New York, and Palmer (in a comment to the Rangoon Gazette, quoted in Counter-Offensive, 174) then used this refusal as an indication that Basic was doomed to failure, its exponents unwilling to meet with reasonable men.

According to Bongers (133f), Harold Palmer was invited in 1943 at the behest of Churchill to consider changes in Basic which would make it more useful as an international medium. Palmer suggested the addition of 'an adequate number of verbs' to those already in use, so that give him a push could safely return to push him. He advocated the addition of more grammatical words and that compounds of a non-

Standard nature be replaced by their Standard equivalents. The number of content words should also be increased. Bongers notes: 'But with these modifications it would no longer be the Basic as planned by Ogden and Richards: it would be a Basic Standard English as envisaged by those various individuals and bodies who have for so many years specialised on the subject.' Bongers favoured such an adaptation.

Ogden is important to lexicology because he attempted to organize a clearcut minimum vocabulary, and to lexicography because this organized vocabulary was put to use in due course in a dictionary which rivals the work of Michael West. In 1940 there appeared The General Basic English Dictionary, 'giving more than 40,000 senses for over 20,000 words', the definitions throughout being framed in Basic. The dictionary was intended for both the young user (strengthening the implication that Basic derives from a special kind of English vocabulary) and the foreign learner. The prefatory note claims that the 850 words of Basic are 'naturally the key words for Dictionary purposes', as a kind of direct reply to West's defining vocabulary. The book is not however restricted to students of Basic; anyone using the dictionary and already having some training in English, 'through Basic or any other system, will be able to make headway by himself with the English of Library, Radio, and Newspaper.' The dictionary claims that the circularity of definition so common in conventional dictionaries is avoided through the medium of Basic, a point which suggests that Ogden has something in common with Cawdrey.

If the West and Ogden approaches were vehement competitors for the allegiance of language teachers, then some check on their comparative merits should be possible by examining respective entries and

definitions in the Ogden and the West dictionaries. To test the differences, something like the following brief comparison might be adopted. I have taken Cawdrey, the earliest similar lexicographer, as a guide, and here give the same four words that I used in the second chapter, when indicating how similar Cawdrey is to the modern Random House Dictionary (1966). The West material is from the International Readers Dictionary (1965) while the Ogden material is from the 1955 Evans edition of the General Basic Dictionary:

	<u>West 1965</u>	<u>Ogden 1955</u>	<u>Cawdrey 1604</u>
<u>edifice</u>	a building	building, sp. great	building
<u>impose</u>	lay a load or unpleasant duty upon someone	put (tax, punishment etc on)	lay upon or put on
<u>prompt</u>	ready, done at once	acting or done quickly, readily	ready, quicke
<u>ruminate</u>	bite food over and over again, like a cow; think	be biting the cud; be turning question over in the mind	to chew over againe, to studie earnestlie upon

The similarities between West and Ogden far outweigh the differences, and this small sample reflects a general similarity which can be seen in the books themselves. Syntactically the continuous tense in Ogden is typical of Basic and one of the flaws assailed by critics. However, both Basic and West have the word bite, presumably through a shared unwillingness to add chew to either minimal vocabulary. If however naturalness of English is the criterion for assessing all three, then Cawdrey from 300 years ago wins handily. More impressive for general lexicological speculation are, however, the over-riding similarities among all three.

Rose Egan kept her strongest censure for synonymists whose methods differed minutely from her own, and so it would appear with West and Ogden. The minimal nature of their differences is borne out by Fries and Traver (73ff) in a very suggestive analysis. In comparing seven lists (Ogden, West, Palmer, Thorndike, Faucett-Maki, Aiken and the Carnegie list which we shall shortly discuss), Fries and Traver found that, whereas there was only a 50% overlap between Thorndike's first 1,000 and Ogden's Basic, between Basic and West the overlap is 79.2%. The overlap between Basic and the West-inspired Carnegie List is the remarkably high 93%, considering that Ogden was not at the New York conference. West's vocabulary has 99 of Ogden's 100 grammatical words and operators, 84 of his 100 qualities, 154 of his 200 picturable words, 291 of his 400 general words and 46 of his antonymous qualities. In every case the later Carnegie list moves nearer to Ogden, not farther away, the most startling concession being to include 369 out of Ogden's 400 general words. Ogden may not have attended the conference, but his words certainly did.

3.6 The Carnegie Report and the General Service List

In October 1934 the Carnegie Conference was held in New York, attended by West, Palmer, Thorndike and Faucett, among others. An initial report was prepared, and was developed at a second meeting in London the following year, when arrangements were made for a tentative 'general' list to appear. This list and the comments with it would serve as an 'interim report', hopefully the basis for future work.

The Conference did not seek total reconciliation or synthesis between the objectivist and subjectivist positions, but was in fact

quite successful in harmonizing them: 'We noted a close correspondence between our judgement and the results of the objective method ... within the first 1,500 words. Beyond the first 1,500 words we found neither the purely subjective nor the purely objective lists satisfactory. We observed that 3,000 words is an unsatisfactory level, since above 2,500 words stylistic variants appear in such numbers that it is difficult to draw the line.'

The group rejected what they called an 'island vocabulary' serving as a self-sufficient simplified language, but as we have seen this did not prevent a closer adaptation of lists towards Ogden. They accepted Palmer's argument that a general list should be supplemented by special lists for specialized purposes: the classroom, agriculture, technology and so on. Such lists were never made, but one can assume that if they had been made they would have begun to show, through the spread of topic headings, a conventional list not unlike Duden and the older cosmic orderings. It is interesting therefore that Duden came into English in the same year as the Carnegie Report was published (1936).

Considering themselves as initiating a period of research into words, the Carnegie Group listed 21 areas worth further study, including compounds, derivatives, idioms, sense differentiation and cognate vocabularies (such as Latinate material in both French and English). As Bright and McGregor point out, however, the Conference was 'not at the beginning of a period of work on vocabulary but nearly at the end' (22). It is clear from the report that certain problems stood in the mid-1930s pretty much where they had been for decades: compounding and derivation and idioms were as much in need of clarification as ever, and homonymy, polysemy and synonymy were as hard as ever to elucidate.

It is also interesting that, despite efforts to present a united scientific front, the group exhibited features identifiable with past etymologists and lexicographers. In Stoic fashion, they regretted 'a tendency at the present towards a loss of stylistic values', and assumed that 'a selected vocabulary offers an opportunity of purging the language of words and other items which tend to be misused'. This point of view is markedly similar to the third point in Ogden's claim for Basic and is reminiscent of moments in Roget, Webster and Johnson. The members also took a special view of sex roles in language, as pointed out by Bright and McGregor (21), where they note with regard to the names of tools: 'Our criterion was that the tool should be such as might be found in any home and used by women as well as men. Thus we may include Saw and Hammer, but exclude Drill.'

A general list of some 2,000 items emerged, and has served since as 'the accepted and indeed the only objective authority on what to include in syllabuses for the teaching of L2 English' (Widdowson, 123). Bright and McGregor (22) consider that the list has 'stood the test of time remarkably well.' It was developed in due course by West himself into its current and probably final form, A General Service List of English Words (Longman, 1953). G.B. Jeffrey in his preface rather diplomatically wrote an epitaph to word-listing when he observed:

'As a member of the distinguished band of workers in the field of vocabulary selection who attended the New York Conference in 1934, and one who has continued to interest himself in the subject, it is fitting that he (= West) should bring this phase of research to what looks like a culmination, since attention is now shifting to structural problems' (vi).

The General Service List (GSL) is the acme of achievement in the movement, and as such it is worth comparing an entry from it with an equivalent entry in Thorndike and Lorge's 1944 compilation. The entry for game in the latter is:

GAME AA 700 966 700⁺ 639

whose rather cryptic symbols can be explained as follows: AA indicates that game occurs over 100 times per million running words, the 700 referring to appearances of its token in the pure Thorndike of 1931, the 966 referring to frequency in the Lorge Magazine Count, the second 700 to the special Thorndike Juvenile Count (the asterisk indicating an estimate rather than a precise figure), and the 639 relating to the Lorge-Thorndike Semantic Count (without details). This can be compared with West, which incorporates details from the Lorge semantic count:

GAME	638	(1) (amusement, children's play)	
		Fun and games	
		It's not serious; it's just a game	9%
		(2) (with the idea of competition, e.g. cards, football, etc)	
		A game of football	
		Indoor games; outdoor games	38%
		(3) (a particular contest)	
		We won, six games to three	
		I played a poor game	
		Playing a losing game (10.5%)	23%
		(4) (games = athletic contest)	
		Olympic Games	8%
		? (= animals, 11%; game-/, game-birds etc., 5%)	
		(= fun, Make fun of, 0.5%)	

For the teacher or other consultant looking for some basic guiding information, West is highly informative, though space-consuming. His 638 refers to frequency in a count of five million running words of the

Thorndike type. In 9% of these occurrences it meant the first sense, and so on through the senses following the Lorge system, the question mark indicating the compiler's doubt as to the worthwhileness of teaching the residue of material in brackets. Minor meanings tend to be omitted in West's articles, and so the percentages tend not to add up to 100. West reminds the user that in studying the frequencies and their percentages, the total frequency of any word should be taken into account, 1% of pull being worth far more than 1% of drag. He also reminds users that it is primarily a list for the written language, and is low on colloquial, stylistic and emotive material.

It has one other serious problem, shared with the slowly garnered NED and the conceptual scheme of Roget. West's list is now dated, a point which I established in a study of a textbook series which overtly claims to draw most of the 700 words of its elementary workbook from the GSL. This is the New Concept English series of L.G. Alexander (Longman, 1967). A check of 12 pages of the vocabulary index in the German teachers' edition, 1969 (for the volume First Things First), shows that out of 354 items (that is, just over half those used) no less than 73 are non-GSL. Twenty items from the 73 newcomers are:

baker	calendar	instalment	licence
beef	climate	intelligence	magazine
beer	dentist	job	measles
blouse	fare	label	mince
butcher	honey	lavatory	mum

The sound pedagogical reasons for adding these items to the list include the need to up-date it for modern school-users, but such massive infusions cast doubt on the on-going status of the GSL as a general list. Without words like these it can hardly serve. Assuming failure to undertake a new frequency count either on the old lines or with some

specific word-theory, any adaptation such as Alexander has made becomes only one more subjective exercise in list-making, to be challenged by the first conflicting 'vocabulary sense' of the first commentator.

One should acknowledge that the Carnegie Report and the GSL stand isolated for criticism in a way not intended by the original Conference members. They worked on an interim report, planning supplements and refinements which never materialized. Circumstances have made their achievements appear monolithic and authoritarian where they probably had no intention of putting them forward as such. In all probability the compilers would concur with later observers in assessing shortcomings, even if they were prone to castigating each other's approaches at the time.

3.7 Commentators on the Lists

I have taken Fries and Traver in the United States and Bongers in Holland as historians of the vocabulary-control movement. This is true in the incidental sense that they made historical comments while engaged in specific tasks. Fries and Traver were asked by the Committee on Modern Languages of the American Council on Education to assess the word lists. Their report is consequently entitled English Word Lists: A Study of their Adaptability for Instruction, and remains a classic of its kind. The thorough-going work of Bongers (Vocabulary Control, 1947) relates to the wish to provide Dutch schools with a 3,000-word vocabulary for their four-year English course. The 'historians' therefore do not stand clinically apart from the movement: in important ways they are part of it.

A comment of Fries and Traver is apposite in any commentary upon the vocabulary controllers: 'One cannot survey the building of word-lists ... without appreciating the immense amount of work that has attended the creation of these lists, as well as the wide practical experience of those who have laboured upon them. It is with diffidence, therefore, that we offer the following conclusions to which we have come from this study' (87). The following is a resumé of their conclusions:

- 1 A restricted list of useful words limited to useful meanings seems not only valuable but necessary in the teaching of a foreign language. In such lists, the status of the 'word' must be clearcut, and figurative extensions should be adequately handled. People tend to forget how fluid meaning is.
- 2 Word lists should be compiled primarily to obtain the symbols for things and for qualities. In this respect, Ogden's Basic English is essentially sound, with its 600 'things', some picturable, some qualities.
- 3 The separateness of 'operations' from the rest of the vocabulary as in Basic is fundamentally important for foreign learners, and verbs can be classed satisfactorily along with preposition-adverbs and conjunctions for this purpose. The limitation of the number of such operators in a first list seems also a sound approach, especially when we consider how wide a range of meanings the common verbs have. The 'learning weight' of a list with verbs as vocabulary units is considerably greater than that of a list in which the verbs are reduced to a minimum.
- 4 A limited list of words with a narrowed range of senses, along with a basic grammar, is useful for the foreign learner, especially for his productive work. Equally clearly, the list and grammar are valuable for text simplification.

5 Passing from such a limited vocabulary and basic grammar to an understanding of Standard English is beset with problems, and at this point other lists need to be added, influenced by quantitative information about words and their senses. There should be two distinct sides to such lists: the productive and the receptive, but what degree of overlap is likely Fries and Traver did not know.

6 Despite the massive amount of work done, there are areas where more research (guided by linguistic principles) is needed:

- (a) colloquial English, however difficult to obtain (and counters should go to the Sears Roebuck catalogue as willingly as to the Classics).
- (b) lists relating to diverse age levels and social situations.
- (c) more than anything else, quantitative and qualitative information about patterns of derivation and compounding, extensions of meaning, shifts of meaning according to the grammatical function of a word, and the relationship in English between form and function.

Bongers experimented in the early 1940s with the 3,000-word list laid down by Palmer in the second IRET report of 1931. Working in three Dutch schools, he set out to ascertain how many of these words students in fact knew (in presumably at least one sense) at the time of matriculation. He concluded that between 2,500 and 2,800 of them were known through current methods, and this strong correlation encouraged him in the assumption (running counter to the Carnegie Report) that a 3,000-word list was a reasonable goal, and that Palmer's was the best of its kind.

When visiting Japan in 1934 he had been impressed by Palmer's claim that the second IRET list would cover 95% of the words in any

English texts of a general kind. The first 1,000 would cover 85%, the second 7% and the third 3%. This gargantuan diminishing return on effort reminds one of the Carnegie view that once past 2,500 we are in a quagmire, but Bongers kept to 3,000, agreeing that anything larger would be counter-productive. He checked the Palmer claim by analysing the first thousand running words of thirteen English works, and got the following results:

Galsworthy, <u>Swan Song</u>	95.4%
Shaw <u>Doctor's Dilemma</u>	96.1%
Temple Thurston, <u>Sally Bishop</u>	96.9%
Upton Sinclair, <u>The Jungle</u>	95.1%
A. Bennet, <u>The Card</u>	95.2%
Th. Dreiser, <u>An American Tragedy</u>	96.1%
A. Allardyce, <u>Unwillingly to School</u>	96.6%
Sinclair Lewis, <u>Dodsworth</u>	94.8%
H.A. Vachell, <u>Quinney's Adventures</u>	93.8%
H.G. Wells, <u>Mr Bletsworthy on Rampole Island</u>	94%
R. Kipling, <u>The Bridge Builders</u>	94.6%
J. Conrad, <u>Typhoon</u>	91.3%
<u>The Lagoon</u>	95.8%

The success of the Palmer list is impressive. Bongers however wanted to improve on it by examining other lists, arranging their contents on Palmerian principles and standardizing their frequencies. It was an immense undertaking, requiring massive and ingenious manipulation of the lists (cf Bongers, 140ff), resulting derivative lists being labelled according to letters of the alphabet. The lists examined were Palmer's second IRET, the Palmer-Hornby 1,000-word English, Faucett-Maki, the Carnegie General List, Thorndike 1931, and Helen Eaton's Comparative Frequency List for major European Languages (1934). Both Bongers and the Fries and Traver approach indicate the incestuousness of lists by the 1940s, but it is significant that Bongers ignores Ogden in his work. After much re-arrangement, shifting, comparison

and collation, rejection and re-alignment, Bongers emerged with a composite set of K, L and M lists, for the first, second and third thousand words of English. The details of this tour de force can be found in his general study.

Palmer claimed a coverage of 95%, and Bongers advanced this to 97.5% for the KLM 3,000, broken down as 89.5% for the first thousand, 5.6% for the second, and 2.4% for the third. A comparison with Palmer's original claim and breakdown suggests that Bongers may have improved most on the first thousand, where it really matters. It seems from one viewpoint a marginal increase for so much effort, but from another it is considerable (if spread over lengthy texts). Bongers provides comparative tables to show the capacity of the six lists to cope with various specimen texts. None of the lists does really badly, and this is one of the justifications for the whole massive effort since 1900, but the KLM has the following showing, over the first thousand running words of ten texts:

An article in <u>The Times</u>	96%
Grenfell, <u>A Labrador Doctor</u>	96.4%
Pearl Buck, <u>The Good Earth</u>	96.4%
James Hilton, <u>Good-bye, Mr Chips</u>	96.9%
Emily Bronte, <u>Wuthering Heights</u>	97.4%
Warwick Deeping, <u>Old Pybus</u>	97.4%
Shaw, <u>The Intelligent Woman's Guide to Socialism</u>	97.5%
Dickens, <u>A Tale of Two Cities</u>	98.6%
Anna Sewell, <u>Black Beauty</u>	98.8%
Eleanor Dooly, <u>The Radium Woman</u>	99%

Bongers considered that the marginal improvement was worth the effort, and in getting it pushed the collation of lists to its limits. Assuming that his list suffers from aging in the same way as West's, it

remains the *pièce de résistance* of the movement in sheer tenacious effort to reach maximum coverage.

We are now in a position to consider Bongers's own conclusions about vocabulary control. He had a greater commitment to it than Fries and Traver and realized that there was nothing specially new or revolutionary in trying to get at the common words of a language in order to teach them before the rare ones. He thought that word-counts do determine frequency, but have no significance under a million running words, and even then are only the raw material for subjectivists to adjust. Indeed, he denied the reality of objective control, arguing that the selection of the texts for counting is subjective. To handle words in and after a count, the principles of Palmer were in his opinion the best, and those of Ogden were irrelevant because Basic is not really English. If 3,000 is an acceptable upper limit for school courses, then the KLM list is the best provider, and analyses of the extent to which any book conforms to that list should, in his view, be printed in the preface, to help teachers assess the usefulness of the book to any class.

The outstanding difference between Bongers and Fries and Traver (both writing at about the same time) is in their attitudes to Ogden and quantitative studies. The Americans rate Ogden highly and urge further carefully defined quantitative work. The Dutchman dismisses Ogden and suspects quantitative work, except as raw material. Bongers is relatively satisfied with Palmer's word theory, while Fries and Traver survey Palmer but do not include his views in their conclusions. Unlike Bongers but in concurrence with the Carnegie members, Fries and Traver want more research on many aspects of lexicology. While Bongers

shows an awareness of the studies done by Fries and Traver in the 1940s, the Americans on the other hand appear unaware that in the Netherlands he was pushing the minimal-vocabulary concept to its practical limits.

3.8 Conclusion

If we can distinguish within lexicography two associated but distinct traditions, glossary and vocabulary, then the work of the vocabulary-controllers can be linked with both. Where necessary, for example, they alphabetized like glossarists and in the end became dictionary-makers in their own right. At the same time their interest in basic vocabulary for educational purposes rather than hard and odd words puts them on a par with the vocabularists, and, indeed, every time they proposed special lists beyond some primary list they were implying the creation of topic headings. Given more enthusiasm and time, the movement might have emerged with ABC lists within topic schemes. Unfortunately, however, the special lists were never compiled, either quantitatively or subjectively.

The vocabulary-control movement arose out of an interest in how and where we use words: in stenography, spelling, reading and language learning generally. Inevitably this interest expressed itself, as lexicography has always done, in terms of the written medium, and began somewhat crudely (but probably unavoidably) with the counting of graphological tokens. Only in later assessments, when it was virtually too late to start it all again, did it become apparent that a theory of what to count and where to do the counting was a prime necessity. Palmer and Ogden came nearest to a theory of words, but for most members of the movement a rather incestuous juggling with lists appears to have been preferable to going back to first principles.

The failures of listing to date do not necessarily mean that listing should never under any circumstances be attempted again. Mechanical aids undreamt of by the early listers now exist to take the tedium out of the work, and to provide results more rapidly: spoken language can be got at now through recording systems in a way unimaginable even in the 1930s. Whatever work might be done in future however should be linked with a model of the lexicon and cannot be seriously undertaken until the following areas are adequately delineated:

- 1 sense differentiation (both homonymy and polysemy)
- 2 the relation of derived words to the forms from which they are derived
- 3 the relation of compounds to the forms from which they are compounded
- 4 the question of figurative extensions of meaning in any item
- 5 the relation of words one to the other in Palmer's 'constellations'

These are the crucial problem areas in lexicology. Beyond them lie further problems of stylistic variation, diachronic change and geographic differences, but a solution even in part to such questions could help resuscitate a brave movement that failed for want of a clear methodology.

We have seen that wordbooks of any kind tend to be culture-bound and period-bound, and such expressions as 'limited' or 'dated' pejoratively suggest that we can rise above such considerations in some way. The efforts of West and Bongers suggest, however, as did those of Murray and Roget, that lexicography can never escape period and place, and that in fact we should accept such facts of life. Like any other book, a wordbook is a fossil from a synchronic stratum that recedes into the past. The very vaunted 'newness' of a new book is the surest guarantee of its being dated a decade later. Adequate

opportunity for revision seems therefore necessary in tomorrow's word-books, whatever the tradition they appear to derive from.

Whether objective or subjective or logical, the major lists serve to highlight attitudes to language and aspects of language which are growing more familiar as we proceed. Listers proved to be interventionist like Chesterfield and Johnson, and conscious of deteriorating standards, like the Stoics. At the same time, they undertook their work with a brave new optimism and the almost evangelical belief that they were offering a means of improving tomorrow's English. Even the most objective, Thorndike, indicated that he expected certain values not to change in the world, and chose the Bible and the Classics as texts to count. However apparently disparate the approaches of objectivist, subjectivist and logician might be, and however acrimonious they might be towards each other, what stands out in the end is the marginal differences among lists, particularly between the two foremost antagonists, West and Ogden. Civil war is always the fiercest, and in this rather ill-starred movement is particularly to be regretted.

4 Linguistics, the Morpheme and the Word

4.1 Introduction

Popular thinking has always linked 'language' and 'words' more closely than, say, 'language' and 'sentences'. And the more literate a community, the more likely are its members to think of 'words' as important and durable entities, things with white space before and after them on the printed page, with their place kept for them in dictionaries, their fixed meanings and their special histories. We have seen that scholars have generally set out to compile wordbooks and word-lists on the assumption that we all know what words are. In a general sense this seems to be so, but this does not mean that we can just pick them up and categorize them when we feel so inclined. The plight of the word-counters demonstrates this very clearly.

Aware of such problems, modern linguistics has appeared to seek release from the tyranny of words. It has turned to a variety of other theoretical units to provide safer foundations for grammatical theory, but it has not carried the mass of educated language-users with it. Teachers and writers still distress linguists by assuming that the primary aim of their discipline must be the sorting out of words. This is not wholly surprising. The word has a venerable history in etymology and philology, and, as I hope to show, it has also proved pretty indestructible in modern linguistics.

4.2 The Three Ages of Etymology

The philosophers of ancient Greece were interested in man and his relationship with nature. This interest prompted a study of language,

and particularly an examination of its origins. Philosophers debated whether language was a thing with eternal unchanging principles or whether it was the product of conventions among men. This debate about nature and convention (cf Dinneen, 1967:74ff; Lyons, 1968:4ff) led the Stoics to the view that all human languages are in a slow state of decline from erstwhile perfection. A philosopher, so they thought, should set himself the task of tracing nouns and verbs back to their original etyma or 'true forms'.

The Stoic period can be called the first age of etymology, when people looked back to a glorified past and saw change as decadence. This view of things is by no means dead. It is possible for someone surveying the European vernaculars to regard French and Italian as the wreck of Latin, and for individuals to insist that an ancient meaning is 'true' while a modern variant 'corrupt'. We have seen how strong this view has been in lexicography. Linguists regularly warn against just this kind of 'etymological fallacy'. Warburg (1968: 348ff) warns against the practice of saying things like 'The proper meaning of horrid's really "bristling" -- from horridus, you know.' Robertson and Cassidy (1954 (1971:91f)) draw attention to a manual of 'good' English which insists that dilapidated can only be used of stone buildings, because it comes from Latin lapidem, 'stone'. The fact that Warburg and others feel obliged to name and point out this 'etymological fallacy' is proof of the tenacity of the Stoic view.

Linguists may disagree with the view that languages deteriorate and that words decay, but it is not something which is easily proved or disproved. For lexicology, the persistence of the attitude, the willingness to think in such terms, may be the interesting thing, and

not the eradication of the sin. From a god's-eye view of English and its vocabulary, one meaning of horrid is certainly 'bristling', even if that meaning is out of tune with the late 20th century, and one meaning of dilapidated links it with stones, although in our times wooden huts, cars and people may get into the same kind of state. The fallacy lies less in the etymology than in the failure to see how words extend and change themselves.

Stoic concern with the true nature of things passed via the Romans into the melting pot of Classical and Judaeo-Christian thinking which followed the decline of the Empire. Etymology entered its second age with St Isidore of Seville (d. 636 AD), whose 20-volume work called Etymologies is one of the early European encyclopaedias. Its subject matter ranges from grammar and theology to farming and house-keeping. Isidore deferred to the Greeks, but his interest was not essentially historical: the origin of a word could be reached as well by inspired 'interpretation' as by rummaging in the history books. He was interested in associations (through derivation of one word from another, through similar sounds, through onomatopoeia, logical contraries and the names of people and places). Whether such associations were objectively verifiable was not important.

Isidore's method of association (such as homo, 'man' with humo, 'from the soil', because God made man from the earth) was a homiletic aide-memoire for students learning special vocabularies. Their language teacher, Isidore, was using the aide-memoires to din the words into their heads. Viewed in this light, even his egregious derivation of cadaver ('corpse') from CARo DATA VERmibus ('flesh given to the worms') has some mnemonic value.

His work can be seen as a systematic version of what is nowadays called 'folk etymology' -- the ordinary thoughtful person's interest in and speculation about where a word began, usually in an effort to get at its meaning. This kind of distortion is said by purists to lie at the root of, say, the current humble pie rather than the older umble (= entrail) pie. Similarly, uproar might well be associated rather with roar than the German aufbruch ('upsurge') from which it came in the 16th-century translation of Luther into English. Folk etymology may be condemned as illiterate (or worse, quaint), but its interest may lie less in the need to condemn it and more in the need to account for it as part of the sociology of language. Isidore has his followers still.

He belonged to a cultural climate now largely gone. His universe was Bible-centred, and the world of most present-day scholars is not. He believed that Adam and Eve spoke Hebrew in the Garden of Eden, and such a belief would immediately condition any inquiry into origins -- as it continued to condition such inquiries up to the late 18th century. His belief in the Fall of Man is sympathetic towards the Stoic idea of language decline, and a literal acceptance of the Tower of Babel could only reinforce such a belief. Isidore worked within the frame of reference of his time, much as present-day scholars work within theirs, and should not be pilloried for it. Etymology also should not be pilloried for having him in its pedigree, any more than modern chemistry would be called to account for the doings of the alchemists. His view of things was still central to speculation about the origins of words in the 18th and early 19th centuries and found expression at that time in the influential theories of the English etymologist, James Horne Tooke, and the American lexicographer Noah Webster.

The third age of etymology came with 19th-century philology. When scholars began (with Sir William Jones's prompting in 1786) to compare Sanskrit with Latin and Lithuanian, the idea of the durability of words despite time and distance was added to the repertoire of etymology. The metaphor under which the philologists worked was, like Roget's, suggestive of trees. It was a picture of language growth, containing families and parents, roots and stems -- what is sometimes called 'the genealogical method of classification' (cf Crystal 1971:152). An Indo-European parent language was conceived as preceding daughter languages like Sanskrit and Latin, which in turn were mothers to Hindi and Italian etc. The Darwinian metaphor of the survival of the fittest in a system of natural selection could easily be grafted onto the original plant, with its concomitant idea of progress, and an element quite contrary to the Stoics was introduced. This suggested that far from deteriorating, languages were improving in the jostle of the centuries, and today's word is better than its predecessor. Such a view is not yet labelled an 'etymological fallacy', but it is just as difficult of proof or disproof as the view that all languages decline from previous excellence.

Edward Sapir (1921) tried to step aside from the issue of the betterment or decadence of any language moving through time. He preferred the metaphors of 'drift' and 'slope', arguing that when a language changes it has a direction (neither good nor bad) and that its users are largely unaware of the slope on which they gently move. Nevertheless, some idea of the slope and the drift can be got by examining the hesitations native users have about the 'proper' use of forms. In Dinneen's words (233), 'This drift is not haphazard, nor the consequence of a lack of system, but is, rather, the stable

dynamism of a developing, living use of language, a factor discussed as analogy in the nineteenth century.'

The question of whether today is worse or better than yesterday is a sociological one, relating to people's situations and expectations about life. It is worth discussion because such attitudes influence language, but whatever achievements etymology may have chalked up, they are independent of assessment on such grounds.

4.3 Skeat: The Root of the Matter

Walter William Skeat is the central figure in the etymology of English in the last century. He took the word as his primitive concept, but as a means of limiting the field of observation (1882:7) and not because he conceived of it in isolation. He considered it mistaken to examine words in modern English without considering changes wrought in them 'in consequence of their grammatical relation to each other in the sentence.'

He emphasized the mixed origins of modern English vocabulary.

In his view, 'no other language was ever composed of such numerous and such diverse elements', and he divided the originating areas into two:

- 1 the 'native' element (Angle-Saxon and the related Scandinavian dialects) as the primary source
- 2 Latin and its derivatives as the secondary source.

His basic division stands virtually unquestioned today. The native-foreign dichotomy serves as the basis for Marchand's study of English word-formation (1960 and 1969), and Skeat's view is recapitulated by Crystal (1971:157) when he says: 'English may be Germanic in one

sense, but is Romance in another, especially when we consider it from the point of view of vocabulary.'

Skeat, then, proposed two distinct source areas, and implied that a contrast still exists between them in the language, but did not examine the contrast to see whether it reflected any distinctions in use or attitude among the users. We have already seen that dictionary-makers have been aware of this contrast for some 300 years.

In his theory of words, Skeat sees a series of historical processes ('consonant-shifting', 'mutation of vowels', 'vowel insertion' etc) operating on a 'root'. This quasi-botanical term is universally accepted today in discussions of word-structure etc., but we should not forget that for the etymologists it is primarily time-linked. Skeat's use was unequivocal: a root is 'the original monosyllabic element which remains after the word has been stripped of everything of the nature of prefixes and suffixes' (280). In his view, all words of Indo-European origin can be traced to roots, members of a set of ultimate monosyllables, and a comparison of the different Indo-European languages enables the etymologist to determine, at least approximately, the oldest form of the root. Such roots are either verbal or nominal in function, the term 'nominal' subsuming both nouns and adjectives.

He does not minimize the difficulties in detecting roots. In his etymological dictionary (1882) he provides a number of appendices to help the student in such work, including a list of the prefixes of English, a statement about suffixes, and select lists of Latin and Greek words of importance in forming English words. His Classical pre-disposition creates a lack of symmetry here, insofar as the Anglo-Saxon

source area is neglected in favour of the Latin source area. More asymmetrical than this, however, is his failure to list the suffixes of English. He argues that 'the number of suffixes in modern English is so great, and the form of several, especially in words derived through the French from Latin, are so variable, that an attempt to exhibit them all would tend to confusion' (630). Very little effort has been made since Skeat's time to classify the suffixes of English as a system of contrasting and functioning elements, and his pessimistic statement about historical fuzziness in the state of English suffixes may be partly responsible for blunting the investigation.

Skeat's historical bias should be noted. He condemns failure in an etymologist to identify suffixes in terms of provenance rather than current function. For example, he criticizes Haldemann (1865) for spoiling a good account of English affixes by misdividing such words as logic and civic into log-ic and civ-ic. 'The truth is' that because civi-c derives from Latin ciuicus it must therefore be regarded as consisting of ciui-, the Latin declensional stem (that is, a root plus its thematic or connecting vowel) and the suffix -cus. 'Of course, words in -i-c are so numerous that -ic has come to be regarded as a suffix at the present day, so that we do not hesitate to form Volta-ic as an adjective of Volta: but this is an English misuse, not Latin etymology' (631). If such is his view of suffixation, then clearly his unwillingness to attempt listing suffixes is not pessimism with regard to current systemic use of suffixation in English, but despair at disentangling the fossil forms of ancient suffixes in the mass of present-day vocabulary. And these are two very different things.

Skeat belonged to the third age of etymology, but a Stoic inclination is clear enough in his remarks about -ic. He faults Haldemann

for failing to apply Latin criteria to English, while accepting that English functions as English and not Latin. Such functioning is for him a decadent misuse of Latin. Truth lies in the past.

He is right of course about the Latin, and Haldemann is right about the English. It might be better to sidestep an unnecessary dispute and ask three questions about the suffix, for which very different answers will be needed:

- 1 What was the use of -i-c in Classical Latin?
- 2 What is the use of -ic among current users of English?
- 3 What historical conditions converted one into the other?

'The truth' would then be different for different periods. If someone chose to complain that the ancient Latin use was more beautiful or pure or efficient, it would be because of a Stoic view of things; if he argued that the modern use is more efficient, concise, scientific etc., it would arise from a kind of Darwinian view; and if he speculated about the data through insufficient information and came up with some clever suppositions, he would be following St Isidore. Meanwhile, of course, people would go on using -ic according to their lights as though nothing was happening in the charmed circle of the etymologists.

4.4 Historical Processes and Grammatical Processes

Skeat was interested in processes acting upon roots and affixes over long periods of time. Traditionally, grammarians in Europe have been interested in process too, but of a rather different kind. From the time of Dionysius Thrax, through Donatus and the Middle Ages and into the 20th century, grammarians have been concerned to show how 'words'

fit together in jigsaws of phrase, clause and sentence. The earliest process models were the paradigms of Greek nouns, adjectives and verbs, transferred in due course to Latin and to the later European vernaculars.

A statement of grammatical process in a paradigm (as for example the cases of nouns like rex, regis, regem etc., or the conjugation of verbs like amo, amas, amat) assumes a prior theoretical form from which others depend. In nouns, this form is the nominative or 'naming' case, and in verbs the first person singular present indicative active is typically used. Other inflected forms are then presented as dependent or posterior.

The desirability of such a process-orientated grammar is not under discussion here. It suffices however to point out that such processes have been used and widely disseminated in language-teaching. They are clearly quite different from Skeat's processes, but confusion has sometimes arisen between them.

This confusion may relate to the use in both kinds of process of the term 'root'. Skeat was interested in stripping away the accretions of time to get at a primordial root; grammarians are interested in 'roots' and 'stems' to which various affixes usefully attach themselves to provide inflections and derivations. Roots are then said to underlie complex forms functioning in the living language.

Using the terminology of Ferdinand de Saussure (1916), we can say that an etymologist is concerned with diachronic developments (through time), and the grammarian is concerned with synchronic occurrences (at the time of use). To avoid confusion between kinds of process, we

can keep the term 'root' with its botanical overtones for diachronic etymology, and adopt the term 'base' for synchronic grammar. This term is not novel (cf Hill, 1958) and suggests a structural or formational minimum on which controlled or predictable building takes place. Given a root-and-base distinction, four theoretical possibilities emerge for the analysis of complex words into their elements:

diachronic root	synchronic base	
+	+	identical
+	-	root only
-	+	base only
-	-	neither

The possibilities can be realized. In the items manly and re-build, for example, man and build are both root and base. In the examples conceive and retain, however, the historical roots are -oeive and -tain, while the synchronic bases for further accretion are the whole items. Thus, the complex inconceivable is synchronically analysable only as 'in-conceiv-able', but diachronically analysable as 'in-con-oeiv-able'.

It follows from an adoption of the root-and-base distinction that if a base is not identical to a root, it must be larger: it must 'contain' the root. Some extra substance (phonic or graphic or both) needs to be accounted for - but usually only by the etymologist, not the grammarian or general linguist.

It also follows that a root may or may not have a meaning in current English, but that a base must have a meaning. The meaning will be nominal, adjectival or verbal, and the force of affixes can be added to it. Put another way, this means that roots are bases from yesterday's language, or from other languages. Today's bases could

become tomorrow's roots, and an ingenious user might breathe life back into one of yesterday's bases for a shorter or a longer time. There is even the possibility that on some occasions in the 'same' word an element may serve as a base, and on others be relegated to the status of a root. A predilection for synchronic studies prevents a proper consideration of such matters, just as much as a prejudice in favour of diachronic studies. We need binocular vision, as it were, to deal best with these problems.

4.5 Lexical Word and Grammatical Word

Etymologists like Skeat had no problem over what constituted a 'word'. It was something isolable from the stream of speech, with space on each side of it on paper and classifiable as a traditional part of speech. More importantly, the words of real interest were nominal or verbal, and their ultimate 'true' forms were the roots. Henry Sweet, in his New English Grammar (1891) took a similar historical view of his subject, but clarified the concept 'word' by distinguishing two categories: 'full words' such as nouns, adjectives and verbs, and 'form words' such as the particles, conjunctions, pronouns etc.

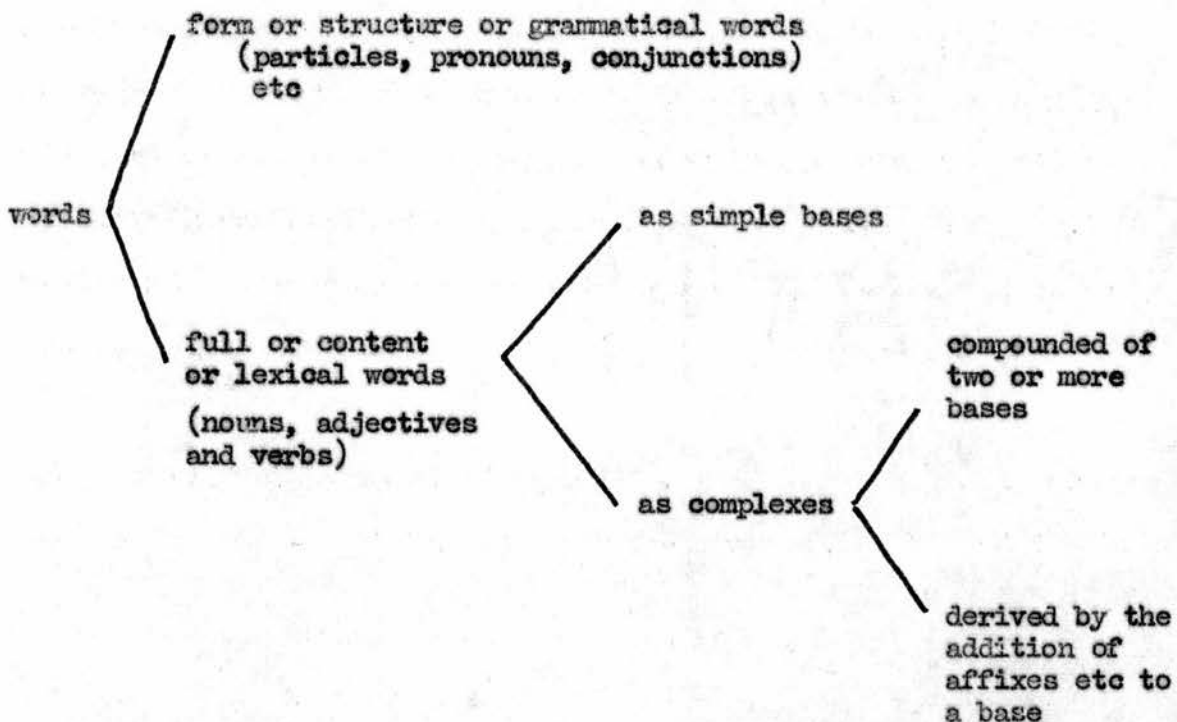
This dichotomy has been followed consciously or of necessity by the majority of grammarians since. Lyons (1968:435) makes a distinction between 'lexical items' on the one hand, and 'grammatical items' on the other, while among the Bloomfieldians, Bolinger (1968:56) distinguishes 'source morphemes' from 'system morphemes'. A dichotomy of 'content words' and 'structure words' is commonplace, and the distinction can be found elsewhere under various names.

It has been asked from time to time whether the form words can rightly claim status as 'words' at all, functioning as they do in much the same way as prefixes, suffixes etc. This kind of thinking led Vendryes (1923) to argue that in language there are only semantemes (full, content or lexical material) and morphemes (processes acting on semantemes), the distinction between a morpheme on its own (a particle or a pronoun) and a morpheme of affixation or accentuation being purely an illusion.

Sapir (1921:33) took a similar line. He considered that words as such are language-specific, not language-universal. His training in the classical European tradition was followed by work among the non-literate Amerindian languages, which demanded a re-appraisal of all traditional terms. His comparison of European and Amerindian language structures however produced the conclusion that only two elements operate in language: radical elements (= our roots and bases) and grammatical elements (= everything else). Certain combinations (including zero grammatical element) in a particular language will constitute its 'words', and these combinations offer the user a unique psychological reassurance. The words of that language cannot be 'cut into without a disturbance of meaning, one or other or both of the severed parts remaining as a helpless waif on our hands' (35). Sapir proposed certain simple formulas to cope with the language-specific word, to which we shall return in chapter 5.

If we assume that English has its language-specific 'words', and also that it is simpler to treat the observable form-cum-structural-cum-grammatical elements as words of a kind because they are written

with spaces between them, then it may be useful to keep Sweet's contrast. A taxonomy can be made for it:



Having made this distinction, we can then (if we wish) discard the form or structure or grammatical words in favour of concentrating on the lexical words. This in effect is what Skeat, Sweet, Vendryes and Sapir are advocating. To their advocacy we can add the observation of others (cf Robins, 1959; Lyons, 1968) that the lexical words belong to a large open-ended class, while the grammatical material, like the affixes, belongs to a class of small closed classes (the pronoun system, the particles of direction and location etc.). Lexicology is primarily concerned with the study of this large, open-ended class, the traditional 'vocabulary' of the language, its 'lexis' or 'lexicon'. It turns its attention to the grammatical class only when they have something valuable to offer in clarifying the nature and function of lexical material.

4.6 The Problem of the Morpheme

The inflectional paradigms which dominated the traditional description of Greco-Latin grammar, and which have been used extensively in the description of European languages, imply, as we have said, the priority of certain forms. In the declension of nouns, for example, the nominative case is taken as 'upright' and prior, while all the other cases are 'oblique', regarded as falling away from the nominative. The etymology of the term case (Latin, casus: 'fall') enshrines this view, a convenient process model for nouns which passes 'roots' through specific 'paradigms' (arbitrarily numbered in Latin grammar as first, second declension and so on) to create a set of specific word-forms in a pre-ordained order.

At the turn of the century Baudouin de Courtenay, a student of the Slav languages, rejected this approach, and argued that in strictly descriptive terms no one case has priority over any other. 'It must not be said that a form of the word serves as the origin for all the rest and becomes another form. The various forms of a word are not formed from each other, but merely co-exist. There naturally arises between them a mental bond, so that they determine each other and invoke each other through association' (in Zirmunskij, 1966:76f).

De Courtenay disavowed both historical and grammatical processes and concerned himself with establishing minimal 'atoms' of language. He used the term phoneme (invented in the 1870s by the French phonetician Dufriche Desgenette) as the theoretical atom in phonological description, and was one of the first to use the analogous term morpheme as a grammatical atom.

As mentioned above, Vendryes (1923, but already in practice by 1914) used the same term morpheme as a process term, for a set of grammatical changes affecting a basic semanteme (element of content). De Courtenay's use of morpheme is radically different and developed its own following in Slavist linguistics, a crucial definition of its role being provided in the terminological glossary of the Prague Circle (TOLP 4, 1931:321), where it is defined as: 'Unité morphologique non-susceptible d'être divisée en unités plus petites, c'est-à-dire une partie de mot qui, dans toute une série de mots, se présente avec la même fonction formelle et qui n'est pas susceptible d'être divisée en parties plus petites possédant cette qualité.'

This morpheme is an element of language as an artifact, a finished product, an arrangement and not a process. It is also not a Saussurean linguistic sign, because it is established formally and distributionally like a phoneme and is not linked with meaning or concepts. It is, however, offered as 'part of a word, in a whole series of words', a statement which echoes down the decades of structural linguistics. The 'word' however is not defined in the Prague glossary. It is not an element of the Prague theory, but rather a sop to general usage, a casual explanatory device. As a result, however, of this casual relation of morpheme-to-word (the contained to the containing) certain tensions were set up in non-process grammatical theory, tensions intensified by the continued use in French linguistics of the term 'morpheme' in Vendryes' sense.

The work of the American linguist Sapir was, as we have seen, process-orientated, and allowed for language-specific 'words' with their own important psychological value. These words were made up,

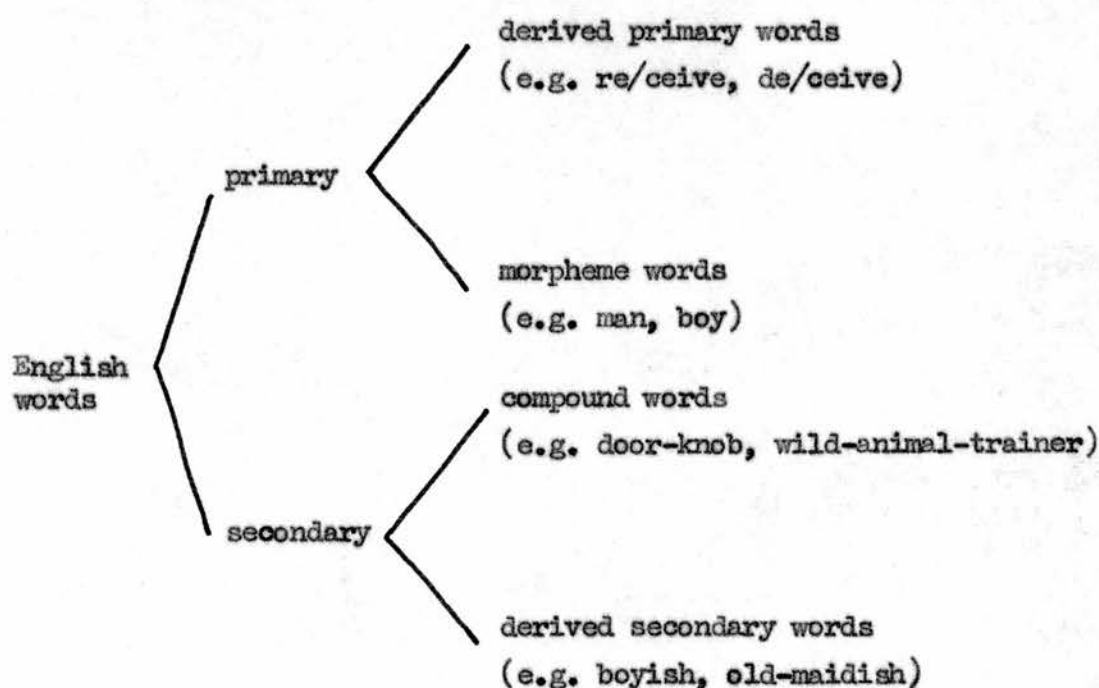
like those of Vendryes, of two distinct elements, radical elements and grammatical elements. Sapir's contemporary Leonard Bloomfield, however, preferred the Slavist view, especially in looking at Amerindian languages as artifacts without a recorded past. His interest in behaviorist psychology also inclined him towards an external arrangement approach rather than a process approach which risked 'mentalist' assumptions. Although educated in the process tradition, Bloomfield rejected the word as a serious theoretical unit (allowing it some status as a 'minimal free form' in such languages as English), and concentrated on the morpheme. For Bloomfield, however, the morpheme is a Saussurean linguistic sign, 'a linguistic form which bears no partial phonetic-semantic resemblance to any other form' (1933a:161).

We are now by this stage discussing three distinct uses of the one term 'morpheme'.

Bloomfield mentions the semantic aspect of the morpheme, but was not enthusiastic about developing it; in practice he followed Prague, being concerned with sound and form, and considering meaning the 'weak point' in language study (140). His morpheme is a bridge for asymmetrical relations between sound and meaning: on the one side it may consist of one or more phonemes, but on the other it can carry only one 'sememe'. Insofar as his linguistics had a semantic side, meaning was additive, a succession of sememes attached to clusters of phonemes, each contributing its mite to the total meaning of any stretch of language. His central concern was to take any such stretch of language and analyse it into its 'immediate constituents' (essentially, its cohering lesser structures such as phrases and morpheme complexes) and then into its 'ultimate constituents', the morphemes. A word of

English is a morpheme complex (which might consist in some instances of one morpheme, or two or three or more), but such words were inherently disagreeable, a mass of irregularities appended to the grammar, so that 'features of selection (in word structure) minutely and often whimsically limit the constituents that may be united in a complex form' (207). Despite the whimsicality however, Bloomfield accepted a ranking of construction within a complex form, inflections being outermost, derivatives next and a kernel, root or base at the centre. Bloomfield talks of such kernels as 'underlying forms', and implies through his terminology that there is a process model somewhere beyond his arrangement. It was probably salutary in the 1930s that attention should so forcefully have been centred in American linguistics on artifacts, end-products rather than formational processes, but Bloomfield's approach had the drawback of being only partly committed to a rigorous arrangement model.

This can be seen in certain inconsistencies in his classification of English words. The taxonomy which follows tries to catch Bloomfield's approach, for direct comparison with Sweet's as shown above (4.5):



The classification shows rather starkly how Bloomfield carried over from his process heritage certain terms and concepts out of place in his new behaviorist arrangement model. The taxonomy is not unlike Sweet's, but omits any dichotomy of 'form' as against 'full' words. 'Primary' and 'secondary' contrasts simple as against complex forms, but the introduction of 'derived' into the description raises serious problems. In the first instance, derived primary words such as re/ceive indicate that Bloomfield wanted to tackle the problem of composites which cannot be analysed in synchronic terms. In the secondary derived words, Bloomfield wanted to tackle grammatical growth of complex from simple. The two uses of 'derived' are process-related, the first to etymology and the second to traditional grammar. They should not have been confused, and even if sharply distinguished, neither should appear in a non-process classification.

Bloomfield had an extensive knowledge of language history (as Professor of Germanic Philology at Chicago), and in dealing with the vocabulary of English in a paper published in the same year as the above material (1933b), he did not introduce morphemes at all. He accepted the traditional dichotomy in English vocabulary as:

- 1 normal or native words
- 2 learned, or semi-foreign words

which is a notable development from Skeat. He noted that affixes of one stream tended to relate to roots of the same stream, and by using the terms 'normal' and 'learned' made a social contrast which Skeat did not attempt. In this, he implicitly accepts the lexicographic tradition which began by contrasting 'vulgar' and 'hard' words. He also indicated that native users of English are not always entirely at home in using the learned, semi-foreign stream. The foreign material is carried in

the written rather than the spoken medium and is consequently open to varied interpretation (as for stress in révocable or revócalbe) and even aberrant formation (as in normalcy for normality). In this study his observations are sharp, and it is worthwhile to contrast the paper with his major work. In the paper, he discusses problems, such as suffix replacement (in forming the noun vacuity from the adjective vacuous), which were to prove insoluble through arrangement morphemics.

4.7 American Structuralism

Many linguists followed Bloomfield in attempting to create a rigorous arrangement model of languages, based on the universal phoneme and morpheme. In this, they extended the analogy between the two units to the following proportion:

phone: allophone: phoneme :: morph: allomorph: morpheme

This extension, essentially achieved by Zellig Harris and Charles Hockett, was considered a major breakthrough and begat many further analogues in the -eme family. The first element in each set is actual language data, an atomic particle that can be observed; the second is also observable, but expresses variant atoms subsumed under the third, the abstract or generic unit. Hence, any morpheme inventory, on the analogy of the phoneme inventory, should state the generic morpheme and the allomorphs which belong to or realize it.

Like anatomists dissecting bodies, the structuralists became concerned with cutting up language data. In describing for fieldworkers the necessary procedures for analysing aboriginal languages, however, the writers of manuals (such as Eugene Nida, Morphology, 1949) found

themselves reluctantly admitting process expressions into their model. Expressions like 'replacive morphemes' and 'stem formatives' appear, and Nida's appeals to meaning as a last resort were fiercely criticized as unscientific by Trager (1951:27). Nida's is an impressive and practical study, but at the very start he defined his subject as 'the study of morphemes and their arrangements in forming words'. In a footnote, he added: 'We are using "word" in this chapter in the usual traditional sense', a point echoing Prague. He neglected however to define this traditional sense of word, and to incorporate it into his theory, and in later pages took over intact Bloomfield's definition of the morpheme.

Within the structuralist movement, the most cogent critics of morphemic analysis were Bolinger and Hockett, the first concerned with problems of meaning, the second with the struggle between arrangement and process. Bolinger (1948 et seq) attacked the diachronic legacy implicit in Bloomfield in the segmentation of such words as away into a + way, disease into dis + ease. He also condemned the atomization of meaning into additive sememes, contending that meaning is cumulative and more than just the sum of parts. Hockett (1954) was radical enough to reinstate process entirely, but the assault was not maintained. He listed three models of description in linguistics, calling two of them 'item-and-arrangement' (IA) and 'item-and-process' (IP), while the third was the traditional 'word-and-paradigm' (WP). He did not discuss WP and probably would have had little to say in its favour if he had, but considered it a defect of his paper that he failed to discuss it. It is significant however that he saw two models in terms of 'items' (a safe, neutral term) and a third, somewhat pejoratively, in terms of 'words'.

The use of 'replacives' and other process terms in IA handbooks suggested to Hockett that structuralists were in fact 'removing the keystone of the whole IA arch', so that 'the model begins to collapse' (394). So-called morphemic elements like replacives were not made up of phonemic material and therefore should not be postulated. However, Hockett saw the possibility that IA and IP could be complementary. Turning to mathematics, he claimed that IA is equivalent to its 'relations' and IP to its 'operations'. A ternary relation like $2 + 3 = 5$ presupposes a binary operation of $2 + 3$, and so the same facts are observable from two positions. Hockett did not however suggest running the models side by side, one for structure and one for formation, but rather tried to offer an IP which would replace IA.

In his process model, Hockett offered to replace 'morpheme' with 'root'. This was not an easy or a happy decision: 'The choice of terminology is difficult here; I do not recommend continued use of "root" and terms stemming from it. "Morpheme" would be preferable, but is avoided here in order better to contrast the models' (396). The transfer to root was however more loaded with problems than Hockett supposed -- because not all morphemes could be roots. Some would be roots, but others, such as inflectional affixes, would be markers of processes operating on roots. The egalitarianism of IA was replaced by a binary approach reminiscent of Sapir, Vendryes and Sweet, one class of morpheme being promoted, the other demoted. Effectively, Hockett in 1954 consigned the structuralists' morpheme to limbo, and although lip service has been paid to it in many manuals since, its definitions have been imprecise and often diffident.

In defining the morpheme in his manual, Dwight Bolinger (1968:53) allows source morphemes (lexical) and system morphemes (grammatical),

but to support his case needs the excommunicated word. He calls the morpheme 'the semi-finished material from which words are made' (53), material available as 'broken pieces that some inventive speaker manages to re-fit', 'fragments of all degrees of standardized efficiency and junkyard irregularity... The only thing a morpheme is good for is to be melted down and re-cast as a word.' After more than thirty years of morphemics this is a pessimistic conclusion, and in coming to it Bolinger does not entirely escape the illusion that somehow the structuralists' morpheme can survive. But as with Hockett, not all morphemes can be melted down and re-cast as words. Plural morphemes, for example, cannot. Such bits and pieces of necessary grammatical demarcation need to be taken out of the argument altogether and should not be equated with the bases to which they are attached.

The British linguist Bazell was one of the most trenchant critics of the Bloomfieldian thesis. He rejected the analogy between phonology and morphology as fundamentally 'absurd' (1949:4) and 'very misleading' (1952:36). Allophones and allomorphs are similar insofar as they are contextually predictable, but on no other count, and three cardinal objections can be made to their continued existence:

1 Allophones are 'motivated', created by the phonological system.

A phonetician can predict them, and his results cannot be improved on by a historian of the language. With allomorphs, however, there can be no appeal to synchronic system. No morphologist would be willing to explain why certain allomorphs (such as the plural endings of oxen and dogs) occur in certain environments and not in reverse (that is, *oxes and *dogen). The explanation, whatever it is, is an outcome of history and not conditioned by the nature of any medium. Unlike allophones, allomorphs can be introduced by analogy into positions which

they did not previously occupy, and so belong to the diachronic as opposed to the synchronic axis of language.

2 Allophones are identified as members of the same phoneme through their intrinsic characteristics, and belong to speech, whereas allomorphs are recognised as members of the same morpheme by the functions they serve and belong to the language as a grammatical system. They are therefore not only distinct from allophones in terms of synchrony and diachrony, but also in terms of parole and langue.

3 No Bloomfieldian theory allows the morpheme to overlap the word, and yet the word is not contained in any theory. The distributionally-established morpheme is therefore dependent on a unit which can variously be established on all or any of phonological, graphological, distributional or semantic criteria.

Bazell does not however reject the morpheme. Like Lyons (1968: 183ff), he prefers to abstract it from the level of visible segmentation of written language data and to regard it instead as a distributional 'factor' without shape. Contacts between such factorial morphemes and phonemic forms are indirect, mediated by allomorphs which he prefers to call 'formatives'. 'It is these processes or relations which the old-fashioned paradigm was designed to bring out, and which modern segmentation is designed to obscure'. The morph is detectable, but the morpheme lies somewhere above or below or behind it, a peg on which to hang indirect correlations of sound and meaning, in a process model which bears a closer resemblance to the traditional word-and-paradigm approach than to either Bloomfield's IA or Hockett's new IP.

4.8 Chomsky and the Syntactic Machine

By the late fifties, the pre-occupation of American structural linguistics with arrangement models gave way to a new process model not directly linked with Hockett's proposal. In it, syntax predominates and morphology is virtually by-passed. With Noam Chomsky's Syntactic Structures (1957), American linguistics was turning away from sound and form towards meaning. Such a turn is only implicit in the 1957 Chomsky, however; the drive towards semantics took almost ten years of experimentation and controversy to reveal itself.

Chomsky's transformational-generative syntax (TGS) is a direct development of Bloomfield's immediate constituent analysis, through the string-analysis work of Zellig Harris (1951), although superficially it appears to reject much of Bloomfield. The anti-semantic spirit in which Chomsky was trained co-exists uneasily, however, with two additional areas in which he was interested:

- 1 pre-suppositions of a logical nature about what underlies certain sentence forms, related by Chomsky himself to the 18th-century Port Royal grammarians, and
- 2 a formalism or system of notation derived from the mathematico-logical work of Carnap and Boole, who (like Bishop Wilkins) were interested in well-defined artificial systems handling meaning rather than in ill-defined natural languages.

It is characteristic of Bloomfieldianism that it made no use of pre-suppositions (which were mentalist) and was exclusively interested in the raw data of natural languages. On the surface therefore the Chomskyan approach appears to be revolutionary and counter-Bloomfieldian.

Against this, however, must be placed the similarity between Chomsky's tree diagrams and immediate-constituent branching analyses, and the experimentation of Harris with transformational techniques. The explicit effort of Chomsky to handle syntax without recourse to semantics and latterly his concern to indicate that syntactic structure is independent of semantics also suggest a Bloomfieldian origin.

For Chomsky and his successors, a grammar is an automaton dealing with ideal states of language rather than natural languages plagued with problems of individual misperformance. The grammar machine explicates the class of well-formed sentences which make up any language L (usually taken to be English). It assigns structural descriptions to these sentences and is said in consequence to 'generate' them. This approach, as a highly abstract exercise, can throw points of grammatical interest into sharp relief and demands a high degree of explicitness in the formulation of rules. Insights obtained in this way can, it is hoped, be checked against natural language phenomena, so that a model running in isolation is as valuable to linguistics as to any other science. Care, however, should always be taken to avoid equating the automaton with natural language, or confusing the model with a description of the real thing.

Being a mathematical construct operating in a closed universe, the TGS model explicitly relates to ideal speaker-hearers of L, each one possessing an identical competence to all the others. Anything not defined by the rules is not part of the language and would instantly be recognised as such by the postulated speaker-hearers. Any wrinkles in the system, when related to natural language, can be put down to

- 1 the interim state of the TGS machine
- 2 its theoretical and abstract nature, or
- 3 faulty performance on someone's part.

Such arguments in effect serve to shield the system against criticism, and as Wilks puts it (1972:67): 'Chomsky has formulated the theory of transformational grammar so as to rule out the possibility of disconfirmation'. It certainly appears that in emphasizing the egalitarian competence of ideal speaker-hearers in an ideal, static universe defenders of the theory tend to assume that everyday speaker-hearers of English are also egalitarians in a static universe, a point very difficult to defend, and one which such investigators as the Glettmans (1970) found it impossible to substantiate.

The theory in due course went through several distinct stages, the clearest contrast being between the 1957 and 1965 versions. In 1957, TGS was a syntactic machine with a 'phrase-structure component' on which a 'transformational component' operated. Certain unspecified basic counters were manipulated in noun and verb phrases to produce a set of kernel sentences, which could then be fed into a transformer to get adaptations or 'transforms': such variants as negative, interrogative, passive and other sentences. Baudouin de Courtenay would have rejected such an approach as having the same priorities as the paradigm: putting one type of sentence first, and then mechanically deriving all the others from it. The 1957 model was, nonetheless, a convenient way of describing sentence relationships in terms of a kind of 'nominative' sentence, and Chomsky was engaged, by and large, in making a Sentence-and-Paradigm model.

By 1965, TGS had expanded into a central syntactic component with two layers (a deep structure and a surface structure) in which the earlier components were now seen as 'sub-components'. Flanking this construct were two additional full components, one for phonology and the other for semantics, syntax taking on its traditional appearance as a bridging system between sound and meaning. By asserting, however, that these two new components served only to 'interpret' what goes on in the syntactic component, Chomsky continued to assert the relative autonomy of syntax. The phonological component could 'interpret' end-products on the surface layer of the machine into sounds, while the semantic component was considered to 'interpret' essential deep elements of sentences in the machine into meanings. As an analogue of language, this version suggests that the making of the bridge has priority over the two areas which it is designed to link.

Such a machine, in either version, requires elemental counters to be fed in at some point, so that phrases can combine into sentences. The primacy of the sentence in TGS is illusory, since it implicitly depends on a scale of units below it (phrases and the constituents of phrases, labelled according to their functions). The sentence is a convenient point at which to set up shop, but as the machine becomes more complex, so the sentence is given more and more absolute value. Without material fed into the machine, however, no sentences could be generated, but whether such material was to be 'words' or 'morphemes' was not considered important in the early stages, when experimenters were more concerned with functional labels such as N and V for noun and verb rather than the sets of items for which these functional abstractions stood. They were free, in consequence, to perpetuate all the ambiguities of the past.

When Katz and Postal (1964:6f) argued for deep-level symbols transformed into a surface-level string, the term 'morpheme' was assigned to the deep terminal symbols, and the term 'formative' to the surface terminal symbols. This approach is markedly similar to Bazell's. As with Bazell and Lyons, no semantic considerations need be invoked for these symbols, and no pronunciation need be assigned to any of them. They are transformed in elegant freedom from both meaning and sound. In essence, nonetheless, whether at the surface or deep down (at one or more than one level of deep structure) such symbols are all the same: to call them 'morphemes' at one stage and 'formatives' at another is tautologous, and the distinction was short-lived, though never (to my knowledge) openly disavowed. Latterly (1965: 65), Chomsky proposed instead that only 'lexical formatives' were available for slotting into the machine at the deepest level, anything happening to such lexical formatives after they had begun to move up the transformational ladder being indicated by process markers of one kind or another. In this, the TGS model moves nearer Sapir, Vendryes and Sweet.

Like the Prague Circle, Bloomfield and Nida, however, TGS proves imprecise on 'words'. Katz and Postal did not provide a place in the theory for any such term, but nevertheless brought them in to explain the transformational process. Thus, any surface structure generated by their rules 'brackets' the terminal strings into 'non-overlapping words', to get the 'actual strings of words of which sentences consist'.

In versions since 1965, the lexical formatives are fed into the sentence-generating machine from an inventory or list called the 'lexicon', which is a static totality. No discussion arises as to

whether this inventory can be modelled as clusters or constellations of formatives rather than just as a list, and it might not be too rash to suggest that in this hazy concept of a lexicon we have the classic preference of theorists for the glossary tradition: there is an ultimately True List to be fed into a True Machine, like rounds on the belt of an automatic gun.

The lexicon in the syntactic component is not, however, the only list available in TGS. A matching list, called 'a dictionary' exists across in the semantic ('interpretive') component. In the lexicon, a formative relates to certain syntactic features such as 'animate' and 'countable', in order to be assigned to its proper place in the deep matrix of sentences. Identical semantic features occur however in the dictionary, a condition which Weinreich (1971:314) sees as redundant. Such redundancy in a machine whose prime purpose and justification is economy of expression is a serious drawback. There is no a priori reason why features should be assigned to an autonomous syntax rather than a semantics, and certainly little to be gained by duplicating them.

In the classic TGS theory of semantics (Katz and Fodor, 1963) it is apparent that more than simple reduplication militates against the two lists. The semantic 'dictionary' clashes sharply with the syntactic 'lexicon' in what can be listed as a member. The syntax allows minimal chunks, presumably some kind of content item which when fed through and 'interpreted' by the phonological component will be a lexical base (Sapir's radical element). Such is the 'lexical formative'. The semantics however appears to allow for the traditional 'words' of English, some of which (the example 'colorful') are lexical complexes, not bases. Any complex such as color-ful cannot however be matched

feature for feature with anything in the lexicon, which presumably can only list color (and perhaps full, raising all sorts of diachronic problems for other items like beautiful, hopeful, careful, restful). In the Chomskyan scheme, unless otherwise cast as a synchronic composite, colorful must be a complex generated by the syntactic component from more elemental material.

In effect, TGS has been raised on the same foundation of uncertainty about 'words' as the preceding Bloomfieldian tradition. Whereas Bloomfield sought to escape their tyranny by referring to a smaller unit, Chomsky avoided them by turning to a larger and very traditional unit. Neither however succeeded in avoiding them.

4.9 Halliday: Word and Lexical Item

The Bloomfield-Chomsky tradition can be contrasted with a British school deriving largely from J.R. Firth, and represented today by Michael Halliday and John Sinclair. The major difference between this school and the American structuralist-to-transformationalist school is that it takes meaning as central to any theory of language and proposes distinct 'levels' of grammar and lexis. Additionally, Halliday's approach avoids over-dependence on any one grammatical unit and instead offers a rank scale in which 'word' is included:

```

sentence
  clause
    group/phrase
      word
        morpheme
  
```

With the exception of the morpheme, this rank scale is traditional and can be established on distributional criteria in the level of

grammar, without appeal to semantics. Halliday's morpheme differs from Bloomfield's in that it explicitly relates to 'word' as a theoretical unit established distributionally, and is not made up of phonemes, which are units on a completely separate level of language. His morpheme, however, does share with Bloomfield's its egalitarianism, embracing both bases and the markers operating on them. It is a unit for grammatical analysis, but has the dubious capacity of being realized entirely by phonic or graphic substance when a base (either bound or free) but variously by such substances or through some other way of signalling it (replacement, voicing etc) when not a base. It therefore serves, as Bloomfield's morpheme does, to disguise process.

This is not always immediately apparent. Scott and his colleagues (1970:7f) in their English grammar based on Halliday and Sinclair conceal the problem (as many definers have done) by first describing the morpheme in terms of discernible affixes, such as -er, -ing and -ed. This helps to establish in the student's mind something 'smaller than' a word. They then extend the term to include lexical bases, which may be bound or free morphemes. This fits tidily into the Hallidayan view that any unit on the rank scale can contain one or more of the units immediately below it, but is a clumsy device (conditioned probably by orthography) in that it constantly demands the equation of essentially different elements: bases and the variety of markers of process. After all, a process marker such as an accent shift or vowel change can hardly be described as a unit 'smaller than a word', and a zero morpheme (as in the conversion of noun man to verb man) cannot be described as smaller than anything. Yet it is offered morphemic status widely in works based not only on American structuralism but also on Hallidayan theory (cf Strang, 1962 and 1970, under index 'zero'; Marchand, 1969:359ff).

Halliday assigns no pre-eminent status in his theory to any one unit on the rank scale. He takes the view, however, that in utilitarian terms the 'sentence' is an important unit for making situational statements, while the 'word' is important for making lexical statements. In lexical statements, the word is to be related to 'lexical items', which need not however be words in the grammatical sense. They are units or wholes which carry unique lexical information. We may call this a distinction between 'morphological words' and 'lexical words'. I prefer to keep the term 'word' for both the grammatical form and the lexical unit because they are both

1 traditionally called 'words'

2 in the lexical instance 'word' is more committed than the neutral 'item'.

I would argue that Halliday himself offers a good defence of such a practice when he says (1961:267) that in his theory provision must be made for delicacy of analysis. In a less delicate analysis, 'grammar' and 'lexis' are separate entities; in a more delicate analysis, they begin to run together, and it is possible to regard lexis as simply a very delicate grammar. If this is acceptable, then the term 'word' which Halliday keeps for grammar can also be extended in delicate analysis to lexis. My proviso would be like Halliday's, that when using any terms one should qualify them sufficiently to indicate just what one wants to say. There is no necessary one-to-one correspondence between 'lexical item' and 'word' in Halliday, and there is no necessary correspondence between 'morphological' word and 'lexical' word in my terminology. For example, am, is and were remain grammatical/morphological, while be may be either grammatical/morphological or lexical, depending on circumstances. Again, in an extreme case, the expression never-say-die-ism is a lexical word, consisting of elements

which may be other kinds of words and also lexical in their own right on other occasions, with the exception of -ism, which is always a process marker. Halliday's theory allows for such rank-shifted formations, however temporary they may be, and both his and my scheme will allow adequately for their analysis.

4.10 Conclusion

This review of the development of etymology, philology and linguistics shows that it is only comparatively recently that students of language have begun to worry about the legitimacy or otherwise of the 'word' as a tool for scholars. We have seen in traditional lexicography and in the vocabulary-control movement a tendency to take the word for granted, and, rather surprisingly, we have also found that even in modern schools of linguistics which overtly seek safer units to work with there has still been a powerful inclination to continue taking the word for granted, using it informally to buttress the formal use of such units as the structuralists' morpheme.

Simultaneous with this ambiguous approach to words and morphemes there has been a tendency in structuralism and transformationalism to discuss the processes of word-formation in terms of 'roots' on the one hand and some kind of markers of process (usually called 'formatives') on the other. Even where the structuralist morpheme has been the dominant theoretical unit, this activity has continued. Such an activity has been tied in with another traditional desire to separate elements in a language's lexicon into words with a grammatical function and those somehow more authentic words which have content value, as with nouns and verbs.

The term morpheme, though not without its value in the initial description or cutting-up of language data (such as in Nida's very practical work) does not appear to be very useful in handling the internal functional and formative problems of languages like English, where a Sapirian approach seems to work better. Morphemes in such languages inevitably divide into syntactic material on the one side, a kind of demotion, and content material on the other, a kind of promotion, while many so-called morphemes turn out to have no conceivable phonic realizations whatsoever.

By assessing the role of the word and its competitors in both traditional and modern linguistics we have cleared the ground for a fuller consideration of the uses to which the term 'word' can perhaps be put in lexicology generally, and in English lexicology in particular. Many scholars have dealt explicitly or implicitly with the problem of words - dictionary and thesaurus-compilers, language teachers, compilers of word-lists, traditional grammarians and present-day linguists - and in their various ways have increased our understanding of the lexis of English. They have, in the main, been in agreement about the nature of the problem, often after years of gruelling work on wordbooks or frequency lists or special languages. By and large they have shown that we face five problems:

- 1 defining the term 'word' usefully and then considering how best to define the senses of specific words
- 2 accurately stating a word's function in a flow of language
- 3 assessing the principles according to which word-formation (both by derivation and by compounding) takes place in a language like English
- 4 considering the relationship that a complex or compound word-form may have with the base from which it has been formed

- 5 considering the question of a word's origin, especially in another language, and its later function in its adopted language.

To handle these five problems we need a theoretical framework within which we can come to terms with an element in language so useful that for centuries we have used it without coming to a clearcut decision about what it is.

5 Towards a Theory of the English Word

5.1 Introduction

At the present time there is widespread academic interest in language. Educationists are concerned with the place of language in their curricula, and computer and information scientists are interested in harnessing human-language analogues for use in machine intelligence and in the easy storage and retrieval of many kinds of information. Such an interest requires, however, units of various ranks (cf Halliday, 1961: 251 and 269) with which people can work, and also a terminology within which such units can be discussed. We have seen that to date linguistics does not adequately furnish either of these prerequisites for the rank of the word, however well it may be doing so for the rank of the sentence.

Since many of these interested people (mathematicians, information scientists, librarians, educationists and the like) have the average educated person's belief in 'words' linguists must for their sake come to terms with words, rather than continue to work exclusively with units such as morphemes which only add another layer of suspect complexity to the problems of understanding natural language.

Additionally, if we find a way to use the 'word' satisfactorily, it will make it possible for us to live more contentedly with past generations who did not question the fundamental nature of the materials they used.

In this survey we have found that the term 'word' covers, in English as in other languages influenced by the Greco-Latin tradition,

a whole cluster of concepts. In this chapter I propose to analyse this cluster, and to move on from it towards a theory which is both diachronic and synchronic, with implications for future work in lexicology.

5.2 A Typology of Words

Many scholars have referred to the multiple ambiguity of the term 'word'. Developing particularly points made by Lyons (1968:197f), Lamb (1969:41) and Mathews (1970:108f), we can seek a typology which may make the handling of such inherent ambiguities easier. These ambiguities relate to various levels of language (such as phonology, graphology, morphology and semantics) as well as to the problem of counting words as types or tokens.

We can begin by making a clear division according to medium (following Abercrombie 1967:1ff): whether a word is written or spoken. This division allows us the following:

- 1 the orthographic or graphological word, the form spelt out on paper, in writing or in print, and probably the most powerful image of the 'word' that we have. This literate image can be regarded as primary in shaping the ordinary person's view of words, as illustrated by such questions as 'How do you spell that word?' Such a word is the product of a number of arbitrary decisions made at different times over thousands of years of struggling towards literacy: an alphabet of a certain kind, running in a chosen direction, with letters of a certain shape and size, with conventional spacing between such letters or ways of joining (or not joining) letters, and inscribed

by various means on various selected surfaces. This kind of word is very much an artifact, and is also the recipient of powerful emotive interest, as we see from time to time in arguments about the 'proper' spelling of such items as colour or through.

2 the phonological word, the form which emerges in speech and is heard clearly by the attuned listener, but not necessarily by anyone else. The attuned listener usually means a fluent native adult, but can in many cases mean an adult with the right technical, professional or other knowledge and expectations so as to be able to interpret the area within a flow of sound which constitutes a 'true' word. We often assume the distinctness of phonological words in the stream of speech, as though they were as distinct as graphological words, forgetting that there are no consistent and unambiguous pauses or boundaries in normal speech, and that syllable boundaries do not necessarily honour the grammatical boundaries between 'words'. We also tend to forget that in highly literate societies (or, more particularly, among highly literate individuals in any society) an ability to appreciate phonological words is influenced by a separate but related capacity to use and appreciate graphological words. Effectively, the realizations of 'words' in the graphic and phonic mediums impinge strongly on each other. Many people are not entirely comfortable with a new word until they have seen it securely stated on paper, when they can, as it were, photograph it for future reference. Others can use phonological words very happily without being sure of how they are realized in writing, or vice versa, can write words without being completely sure about how they are pronounced. The impinging therefore, although for most of us operating all the time, is not total. The mediums are, ultimately, distinct from each other.

It becomes clear from this that these two kinds of words are realizations of some third thing, a kind of discarnate word that may be incarnated in either form. We need a label for such a word, and one which might fit the situation is:

3 the morphological word, which can be said to lie behind both the graphological and the phonological word. Thus, such a word as 'big' has a spelt-out realization and a spoken realization, but is independent of both. There is also nothing to stop a word having two or more spelt-out realizations, as with 'theatre' and 'theater'; and it is certainly true that words have as many spoken realizations as there are variations in the dialectal and articulatory possibilities of the language. To that extent the existence of apparently standard ways of spelling and saying words adds to the illusion that words are permanently standardized things. If, however, someone says that English 'theatre' and Spanish 'teatro' are effectively the same word, then we are indicating that there is an inherent sense of flexibility or fuzziness in our working concept of words, even if not in our more dogmatic assertions about them. The morphological word is consequently what we see as being realizable in a medium (including any other medium besides speech and writing, such as flashing lights or drum beats).

Presumably this concept will cover all English words, but a division emerges here too, the same division that we noted in discussing Sweet and Bloomfield: the dichotomy between 'content' words, and 'form' words. Let us then propose:

4 the lexical word, which raises special problems, because lexical words can be and usually are realized variously by special morphological words. Thus, do, does, did and done are four distinct morphological words, with various realizations in speech and in writing.

Most of us would insist, however, that there is a sense in which they are all the same word, and we would call that word do. This is a recognition of a core of content meaning lying behind the ways in which a word is amended to fit into the syntactic flow. Such amendments, of course, are done by process markers such as affixes, or, in the case of common irregular verbs, by suppletion. Suppletion can often obscure a relationship of this kind, as for example the young child or the foreigner who does not see the link between 'go' and its special past-tense form 'went'. Suppletion establishes for us, however, the need for the lexical word, because clearly 'went' is not an autonomous word in the same way that 'go' is. The lexical word is an abstraction of some subtlety, and should not be confused with the lexicographic word, because dictionaries are constrained by tradition to list 'did' as the past of 'do' and 'went' as the past of 'go' and so on. The lexical word, however, can be taken as the prior condition in any paradigm, so that lexical do lies behind do, does, did and done, and lexical go behind go, goes, went and gone. It becomes a matter of personal preference, however, whether one treats 'has gone' or 'has been going' or even 'has been going to go' as single words, or clusters of morphological versions of lexical words.

5 the grammatical word, which belongs to the structure-creating system of the language and serves to link things together. Such words (adverbial and prepositional particles, determiners, pronouns, conjunctions etc), have a high frequency and have their own special semantics, as for example the particles with their semantics of position and direction, space and time. It is normally assumed that the sets of such words are closed to new members, and such observers as Vendryes and Sapir make us uncomfortably aware that they have a kinship with

derivational affixes. The out in 'throw out', for example, has a qualitative similarity to the e- in 'eject', and we tell children that pre- means 'before' and post- means 'after'. This suggests that grammatical words have a kind of honorary membership of the class 'word' by virtue of the way in which a language is used, or perhaps more cogently by virtue of the way in which we have chosen to think of it, to write it down and so on. Certainly, all grammatical words potentially have the capacity to behave like affixes, as in the 'he' of he-man, and the 'yes' of yes-man. On occasion, they can also be lexicalized, as in the famous 'But me no buts' and 'The ayes win; the noes lose.' The borderline between the two categories is hazy and is made hazier still when we find set phrases such as 'on condition that' which can paraphrase 'if'. If is clearly a grammatical word, so what is the status of condition in such a phrase? Many prepositional and other phrases behave holistically in this way and underline the peculiar semantic-syntactic fluidity of language, between the entirely lexical and the entirely grammatical.

While the logical steps establishing these kinds of words are clear, in practice it is not always necessary to follow them. Indeed, it is remarkable to what extent in ordinary usage we can disentangle the nuances without using an overt theoretical device; but it is as well to have such a means available to us, as the time is probably past when we could afford the luxury of being vague about words. The problems of word-counters such as Thorndike, Palmer, West and Bongers in the period from 1920 to 1950 make it clear that there is no point in starting to count words until one knows what one wants to count, and we cannot know what we can count until we have categories in a typology. This brings us to at least two sub-categories of lexical word:

- (a) the onomastic word, which serves to name things, and may be as simple as Smith or as complex as Smithsonian, as transparent as Johnson or as opaque as Saskatchewan, as immediately relevant in all its constituent parts as the White House or as constitutionally irrelevant as Shakespeare (who shook no spears). These names are clearly lexical, in that they take possessive markers and the like, but differ from other content words in that it makes little sense to count their occurrence and to place them in frequency lists although this has indeed been done.
- (b) the multifaceted word, such as bear (when spelt) and wind (when pronounced), variously classed as a homonym, homophone and homograph, and which raises special problems when one starts to count 'words'. Again, early word-counters chose to count all occurrences of 'bear' as one word, whether relating to carrying things or to a big furry animal, and some did or did not (according to their impulses) include 'borne' as a variation of this word. With the typology offered here, however, and with hindsight, we can in future avoid such pitfalls.

The whole question of counting words however leads us on to a difficult last category:

- 6 the statistical word, which may be any of the preceding, as long as one rigorously sticks to one's initial choice. One can simply count words on a page, and state how many there are. In this case one is simply counting items with white space between them, regardless of what they are, but presumably having taken a decision about how to treat hyphenated words and such solid compounds as 'teapot' as opposed to other forms such as 'coffee jug'. It becomes clear that even in this simple exercise a

consistent set of rules is necessary. If we then wish to count only instances of a 'particular word' on a page, then effectively we are counting tokens of a recurring type, and can say that there are X instances of the word Y. We have to be careful, however, and must decide whether we are counting graphological words or morphological words: Are 'does' and 'did' different words or do we count them as tokens of the type 'do'? If we specify our plan and aims, we can count anything we like, and no one will be confused. Inevitably, however, such problems as bear = carry and bear = animal will emerge to add spice to the undertaking, and further subtle rules will be incorporated until what seemed to be a simple exercise becomes surprisingly complex, especially if we become interested in shades of meaning, in words used in a particular sense. Is 'fire' one word, or two, or three or four in these sentences? -

- (a) He lit the fire.
- (b) They fired the house.
- (c) He fired the gun.
- (d) He fired the man.
- (e) 'Fire!' he cried.
- (f) Her dress caught fire.
- (g) Her eyes flashed fire.
- (h) They were afire with passion.

Here we enter into the issue of polysemy, multiplicity of meaning which may at times appear to be simple homonymy of the bear type, but which at others is, evidently, points along a continuum of understanding which brings in all sorts of assumptions and bits of folk-etymology and questions about whether or not part-of-speech differences, idiomatic

and special accretions change the status of an item in a set.

Lexicology is justifiably concerned with all the aspects of word typology just raised, and must seek to create a coherent picture of how they operate. Coherence is not easy, however, in such a shifting kaleidoscope of possibilities, which ultimately may be an attempt to map a large part of the behaviour of the human brain.

Realizations in speech and writing and even other mediums, interactions between different mediums, frequency of occurrence of types and tokens in texts, morphological adaptations in grammar, relationships and fuzziness between lexical words and grammatical words, and between grammatical words and affixes, and in the use of the 'same' word in different context, the definition of words and their senses, the derivation of later words from earlier words and the creating of compounds - these are all grist to the lexicological mill, but a typology such as is here proposed can help clarify the ground so that the mill can work more efficiently.

Such a typology, however imperfect, makes it clear that a serious problem lies at the heart of any analysis of English words. Having found that the lexical word is, as it were, central to the discussion, we recall the defects of morphemic analysis, that all morphemes were equal, but some were more equal than others. Here again we have seen that grammatical words are, in some sense on a par with affixes, in the service they provide - as markers of grammatical process - for lexical words. They are to be placed among Vendryes' morphemes, Sapir's

grammatical elements. In consequence, within the lexical words, we must look for yet another fundamental and possibly universal category which is not a word as such, but is the irreducible content element, Vendryes' *semanteme* and Sapir's radical element.

5.3 Morphology and Syntax

We have noted above, in discussing Skeat and Haldemann (above 4.3) that truth often depends on the questions asked. This maxim also applies in discussing the traditional distinction made in grammar between 'syntax' and 'morphology'. A common explanation (cf Lyons, 194) is that morphology deals with the internal structure of words, and syntax with the rules that combine words in sentences. Such an approach gives the undefined and multiply ambiguous 'word' primacy of place, and suggests a dichotomy in grammar between larger organization and smaller organization. It then poses the problems of where the boundary is to be set up between the larger concerns of syntax and the smaller concerns of morphology, or even as to which is 'more important'. It has, for example, been argued that Bloomfield over-emphasized morphology while Chomsky concentrated on syntax.

Lyons notes that 'etymologically speaking' morphology is 'the study of forms' and syntax the theory of 'putting together', adding that in older books the distinction is between form (morphology) and function (syntax).

Otto Jespersen took a view of the two terms which adheres closely to the older approach, and obviates the necessity for a larger-smaller

competition between the two. He argues (1954:1) for an indissoluble 'grammar' approachable from the perspective of either form or meaning. When approached from form, from the outside looking in, it is 'morphology'; when approached from meaning, from the inside looking out, it is syntax.

Diagrammatically:

form to meaning
outside to inside
 $O \rightarrow I$
morphology

meaning to form
inside to outside
 $I \rightarrow O$
syntax

Morphology is the observer's side of things, while syntax is the creator's side. 'The same grammatical facts may be, and have to be, viewed from both points of view, but the classification and arrangement must be totally different according as the facts are seen from without or within'. The dichotomy is a procedural device in dealing with two ultimately inextricable aspects of grammar. A morphological question about English might be: How do we use the element -en? The answers might include the points that -en occurs in participial adjectives like drunken, but not productively, and also more widely as an inceptive-causative suffix in harden, soften, brighten etc. A syntactic question about English might be: How is causation expressed in verbs? Among the answers might be that three suffixes are used on different kinds of bases, one suffix being -ize, another -ify and a third -en, which is added to a specific subset of originally Anglo-Saxon adjectives like hard, soft and bright.

Looked at from this point of view, it is not unreasonable to say that Bloomfield as an observer of Amerindian languages took a morphological approach, and that Chomsky with his implicit bent towards semantics has taken a syntactic approach to language.

Edward Sapir (1921), having an interest in the language-specific rather than language-universal nature of the word, developed some formulas for expressing the nature of words in different languages. These formulas are morphological, based on two symbols connected with the concatenating plus sign:

A + b

where A is a radical element and b is any adaptation which can be made to it, a grammatical element. It is lower-case because it is dependent upon A. Its position after A is not however intended sequentially. The b element is regarded as posterior in process terms to the A element, but may be pre-posed, post-posed, in-fixed, or it may be a sound-shift, an accent shift, a replacement or an abbreviation. It is not a segment of any kind but a factor or marker contrasting a posterior condition with a prior one, such as feet as against foot, legs as against leg.

The b element has significance only in terms of the A element, and to show this more clearly, Sapir bracketed b, to indicate this incapacity to stand alone:

A + (b)

At this point we can recall the distinction between a historical 'root' and a contemporary 'base' (above, 4.4) Ideally, the term 'base'

suggests a structural or formational minimum on which controlled or predictable building takes place. A root, however, is ideally something which no longer has such a function (or any attendant meaning) in the modern language. Given a root-and-base distinction, three theoretical possibilities emerge:

- 1 base and root identical, as in the man of manly
- 2 root only, as in the -tain of retain
- 3 base only as in retain itself.

Since elements like -tain did have a meaning and function in times past, we can say that roots are yesterday's bases, and that today's bases could become tomorrow's roots. If we adopt these terms, then Sapir is dealing in bases only, and the process markers which attach to these bases.

Sapir considered the above formula adequate for English words, where the Base A can stand alone as a word, but inadequate for Latin where a base needs some marker to integrate it into the sentence. For the Latin type hortus (hort-us) therefore ^{he} ~~be~~ proposed:

(A) + (b)

The b element might be inflectional or derivational. Sapir took Latin as characterized by the boundness of bases, and English as characterized by the freeness of bases. A moment's reflection however reveals that this is not so. The vernacular, largely Anglo-Saxon, vocabulary of English is indeed characterized in the main by the capacity of bases to stand alone, but the Latinate vocabulary is not so characterized. Consider the three forms:

<u>Vernacular</u>		womanly = woman + ly (base independent)
<u>Latinate</u>	1	normal = norm + al (base independent)
	2	feminine = femin + ine (base and suffix interdependent)

Latinate bases in English may or may not follow the Latin formula (A) + (b). Both Sapir's formulas are relevant in current English, and by using them we can make a descriptive statement about English lexis which is not normally made: that a subset of lexical bases in English does not possess the power to stand alone. Such bases have traditionally been described in etymological terms as 'Latin and Greek roots in English', but clearly they are synchronic bases as well. Femin- has just as much nominal significance in current English as woman. We are faced at this point with clear indications that the morphology of English is not monosystemic but is at least disystemic, as indeed one would expect from a consideration of its history.

Linguists have tended to regard Sapir's formula A + (b) as the formula for English words. When a word does not conform to that formula it is labelled 'foreign' or 'semi-foreign'. Marchand, for example, categorizes word-formation in English as 'words formed on a native basis' and 'words formed on a foreign basis' (1969:5f). In Sweet and many later commentators the apparent foreignness of an element often demanded an illogical classification in morphological terms, obscuring rather than clarifying relationships. For example, in the Vernacular compounds one-sided, two-sided and many-sided the initial numbers are not classified as prefixes, but are numeral bases. In the parallel Latinate words unilateral, bilateral and multilateral, however, the initial elements uni-, bi- and multi- are widely represented as prefixes, and not as base elements in Latinate compounds. A double standard emerges in such cases because one etc can be used as free words while uni- etc cannot. Orthography probably has something to do with this tendency, in that a Latinate word tends to be printed solid (unilateral as opposed to uni-lateral).

The awkwardness of the double standard emerges in the treatment often given to Greek-origin compounds in English. It is not unknown for dictionaries and linguistics manuals to take elements like micro- and label them prefixes (cf Marchand, 1960 (1969:175)), while labelling elements like phon(o)- prefixal or suffixal or root depending virtually on where they appear in a word. Such words as microphone, monopoly, phonology stand in danger of being analysed as consisting entirely of prefixes and suffixes, with no base, rather than more usefully as having two bases connected by the thematic or connecting vowel -o-, while unilateral etc have a connecting vowel -i-.

The same analysts observing Latin or Greek would probably not label uni-, mono- etc as prefixes in their languages of origin, but would prefer to see them as bases. It is only when they appear in English that they develop this unsatisfactory morphological opacity, a kind of second-class citizenship.

It should be noted at this stage that Sapir's point about words being language-specific cannot be set aside. If one wishes - as, say, a lexicographer, word-counter, stenographer, language teacher or other professional person - to deal in words as they are generally understood, then such a typology as proposed in section 5.2 will prove useful. If, however, one wishes to transcend the particular and to discuss possible universals of language, or if one wants to describe more usefully the forces at work in word-formation, then that typology is of secondary consideration. Of primary value in such matters is the dichotomy of language elements into 'bases' (elements with lexical meaning) and 'process markers' (elements of structural organization, with distinctive

semantic and syntactic values). In English, some bases happen to be 'words' in terms of the typology, some do not. There is a difference therefore in how we loosely describe the language in terms of traditional attitudes and expectations, and in how its formative processes can be more precisely described, in possible universal terms. Thus, the item womanliness is an English word (attested for example in the RHD), and can be handled via the typology, but for strict linguistic analysis of its constituents we must go beyond its status as a word to consider the relationship of a base 'woman' to the accretions 'ly' and 'ness'. This distinction may not be immediately clear, but it becomes clear in dealing with less transparent words such as femininity. In discussing English we are not accustomed to giving the base 'femin' the same status as the base 'woman'. Ultimately, however, they have an identical status as what I propose to call 'lexical bases'. Whether a lexical base happens to be a free word or not may be one of the less interesting things about it in morphological terms. A typology of the English word can handle woman, womanly, womanliness, feminine and femininity, but it cannot handle 'femin', a relationship which is crucial inside polysystemic English, as well as in relating English to other languages whose word typologies may be somewhat different.

Additionally, it should be noted that certain Vernacular words in English do not conform readily to Sapir's straightforward concatenating formula. Words such as uncouth, mawkish, ruthless and bashful cannot be said to have bases which are free words *couth, *mawk, *ruth and *bash. Thus, although Sapir's statement is largely true for Vernacular English, it is not entirely true, and this lack of absolute application is an

argument for the ultimate importance of lexical bases in the study of word-formation rather than simply 'words'. These examples, however, also complicate the issue further because it is not easy to assign clear meanings to the bases 'couth', 'mawk', 'ruth' and 'bash'. They are lexical, but they are fuzzy, and they do not stand in isolation as autonomously as 'femin', whose meaning is clear. They need their accretions to give them vigour, whereas 'femin' without any affix is ~~still~~, though not a free word, still the vehicle of a distinct concept.

Sapir does, nevertheless, provide a strong foundation on which to build, and from it we can proceed to the suggestions of a present-day etymologist to help raise an edifice on these foundations.

5.4 An Indo-European Word Formula

The formulas offered by Sapir in the 1920s are simple concatenations: one symbol follows the other, but *b* is not to be seen as segmentally subsequent to *A*, unless *b* is a suffix. Sapir's formula is not usable for segmental analysis of a given word; it is rather a statement of the formational principle in the language.

The etymologist Ross has proposed a different formula, one which can be used in post-mortem or surgical analysis of complex words much as Bloomfield wanted to use morphemes. In a note on morphology (1958: 138f) he makes the following generalized formula for Indo-European lexical words:

$$(P) \quad \overset{f}{\underset{=}{F}} \quad (S) \quad (E)$$

where the brackets indicate optional occurrence, the *P* is any number

of prefixes, the R is the root, the S is any number of suffixes, and the E is an 'ending', an inflectional element. Thus, using the formula, the Indo-European *dom-en-q-s or originating (hypothesized) form for the Latin dominus is analysed structurally as:

R S₁ S₂ E

because there is no prefix, the root is dom, the suffixes are first -en- then -q-, and an ending -s gives the masculine singular nominative. Ross gives other examples from a spectrum of IE languages to show the flexibility of the formula in handling not only hypothesized forms, but complexes current in these languages:

R	root alone, as in Latin <u>i</u> (= go!)
R S	root and one suffix, as in the Greek <u>phere</u> (= carry!)
R E	root and ending alone, as in the Latin <u>pedem</u> (= foot, accusative singular)
R S E	root plus suffix plus ending, as in the Sanskrit <u>bharati</u> (= he carries)

For the Indo-European etymologist, any word form within the range of this formula is a lexical word (with a morphological adaptation), regardless of time or place. In a linguistics orientated towards synchronic description, the success of etymology in charting the patterns of IE words (together with the backdrop of evidence supporting the above formula) has not been given the attention it deserves. If this formula fits IE words for all times and places, then it fits present-day English for its lexical and morphological words. Only one conversion is needed in the formula to give it a dual role: R for root can be changed to B for base, and we have a distinction between the diachronic analysis on the one hand, and the synchronic analysis on the other.

Some illustrations show the advantage of such a dualism: The item manly is RS historically and B S synchronically-grammatically. The item re-build is either P R or P B, but the item retain is either P R or B. Historically, the noun retention is P R S for 're + tention', but currently it is B S for 'retent + ion', appropriate allowances being made for phono-graphological changes of the tain/tention type.

The Bloomfieldians allowed etymological analysis covert entry into their synchronic-morphemic approach to the segmentation of words. Bloomfield, for example, allowed away to be 'a + way', and words like disease were taken as 'dis + ease'. Bolinger urged that such items should be treated as wholes, and that investigators be prepared for degrees of wholeness, which sound like a contradiction in terms, until one considers his example of 'bicycle' and then 'three-wheeled bicycle', in which latter phrase the division 'bi + cycle' is no longer strictly applicable.

Jespersen also drew attention to this problem, particularly with regard to the Latinate vocabulary of English. He argued that elements like re- might or might not be productive prefixes in English, and when they are not (as in retain) they should be denied prefixal status. Our approach to this has been to provide a different analysis depending on whether we want to think of the element historically or not. The problem is not, however, clearcut, as this continuum effect shows:

receive	=	get, obtain etc (<u>re-</u> has no independent value, and is not accented)
repeat	=	do, say again etc (<u>re-</u> has some kind of iterative value, while -peat is opaque and re- is not accented)
re-write	=	write again (<u>re-</u> is clearly iterative, and is accented, and may or may not take tonic stress, and is usually written with a hyphen)

There are clearly degrees of analysability and of wholeness, and Bolinger's warning must be taken seriously, even if only to find some analysis which will help us teach non-native users of English when to stress an element and when not to. There is evidence to suggest, for example, that accent (word stress) is still locked on to the relationship of historical root to affix, rather than linked with the relationship of current base to affix. Two-syllable Latinate verbs like retain and conceive have the accent on the second syllable, and tend to have the vowel in the first syllable (the vestigial prefix) weakened to or towards [ə] in many forms of English today.

We need therefore to have some way of approaching bases which are (in diachronic terms) composites of roots plus other roots or roots plus the vestiges of affixes.

5.5 The Phenomenon of Holism

When I initially raised the issue of distinguishing between root and base, the division between past and present was presented as clearcut. Even then, however, I left provision for occasions when the same item in current English might for one usage be base plus additions of some kind, and yet for another usage the same 'base' might be analysable as a nonfunctioning root. This is actually a widespread and fundamental fact in our use of words in English.

Let us take as an example the Latinate noun appearance. For this word we may propose at least two usages:

- 1 His appearance startled me; it was so sudden.
- 2 His appearance startled me; it was so unconventional.

We can take the first to refer to a person's actions, and the second to refer to his clothes and way of wearing them. In the first sentence, it is possible to paraphrase appearance in some such words as 'act of appearing'. This is then an action nominalization of the verb appear, and so we can agree that appear is a base, and -ance a suffix of current interest. In the second sentence we cannot satisfactorily do anything like this. Here we are discussing a state, not an action, and are forced to give appearance some kind of gloss such as 'clothes' or even 'get-up'. Many Latinate nominalizations and compounds of Greek provenance work in the same way, having an analysable and an unanalysable use, or a more analysable and less analysable use. Thus, in the second situation, the whole word appearance functions as a base. Additional evidence is provided by the fact that the first use of appearance relates to further complexes such as non-appearance, disappearance and re-appearance, but the second use cannot produce such expressions.

This phenomenon can be called 'holism', where the sum of the elements is more than or quite distinct from the parts. Holism, however, is not a monolithic phenomenon: it has its nuances, its more-or-less situations, its own typology. If, for example, one looks at the verb appear, then etymologically we can analyse it as P R, to give ad + pear (Latin ad-parere, 'to come forward into sight'). Now this definition is quite definitely only of interest to etymologists, students of Latin and word-enthusiasts, and a root -p(e)ar is not of much current value to anybody. So appear is already in diachronic terms a holism, while its derivative appearance may or may not be a holism in synchronic terms. If we bring in Bolinger's three-wheeled bicycle, we get another short-

term holism, but it is different from 'appearance = clothes'. Bicycle generally means something with two wheels, and tricycle is available for the other phrase. But this does not deny legitimacy to Bolinger's phrase. It is normal English.

It is possible, therefore, to make a four-fold provisional typology of this holistic factor in English lexis:

- 1 phylogenetic holism, where present-day opacity is great or total, but an ancient or alien analysability can usually be got at.

Latinate examples are appear, receive, horrid, dilapidated and religion. Additionally, many onomastic words (place and personal names) fall into this category: Worcester, Whithorn, Mackenzie, Anderson, New York, Saskatchewan and the like. Finally, terms which are analysable in foreign languages (such as hatha yoga) have to be taken as holisms by most people using English. For certain purposes, however, and in certain situations (poetic licence, educational discourses and technical explanations) even the most opaque, ancient or alien of holisms can be resuscitated and rendered transparent. Thus, the word religion is often referred to as having a 'binding' element linked with the same element in ligament and ligature, but that in its case the binding is in a vow of faith. We can also assume that folk etymological factors can work on such forms, so that explanations are offered in order to make the holisms mean something once again, even if such explanations do not tally with the documented facts.

- 2 ontogenetic holism, where the opacity or transparency of a particular word is entirely dependent on the knowledge of the individual. A user may lack the necessary information to analyse a word, as is the case with many users of such technical terms as hepatectomy,

cirrhosis or sclerotic. They swallow them whole. This variety is important in educational terms, and reminds us of Philip Grove's contention in the Language Bar (1950) that many people are not being educated towards an ability to handle the hard words of English. (For some attempts in this direction, see McArthur 1972c:30 and 33ff). The implication here is that inside English there is for many people a mass of holisms, left unanalysed through ignorance, not through their lack of analysability. This failure could be due to many things, including lack of interest, lack of contact with the discipline in which the terms occur, or misfortune in not having been provided with the key to the door. Additionally, the question of fuzziness arises here, because with many users of the language analysability is more or less a vague skill. A person may recognise that 'vasectomy' is connected with cutting something essential to male fertility, without having a clear idea of vas-, ec- or -tomy or how they interrelate.

3 relative holism, where the analysability of a word is dependent on context and usage, so that in one situation analysis is possible and necessary, but in another it makes no sense. We had the example of appearance; we can also consider geology in the phrase 'the geology of Wales' or philosophy in 'my philosophy of life'. Such relativism does not mean that two senses of a word have split off completely from each other. The success of the holism in 'the geology of Wales' is largely dependent on the cognate meaning of geology in 'geology is the study of the structure of the earth'.

4 rhetorical holism, where, for short-term purposes of comparison or comment, sometimes funny or ironical or provocative but usually entirely practical, a user may render a complex form more or less holistic, that complex being normally quite analysable. We had

three-wheeled bicycle, but we might also consider a brown blackbird, a green blackboard, Martian geography or even Scottish Cheddar Cheese. Such short-term uses may of course become long-term uses, if no other more compact means of describing something suggests itself.

The concept of holism allows one to cope with complex words that are usually more than the sum of their parts, or other than their sum. Greater delicacy in extending or subcategorizing the typology can also bring in several special phenomena which have often given linguists headaches:

- 1 In such words as 'butterfly' there is no way in which the parts can be made to help explain the insect referred to. Such a word is as it were the purest kind of holism; there is no advantage to anyone - outside of poetic fantasy - in invoking either 'butter' or 'fly'. This might be called 'deceptive holism', an invitation to folk etymologists to do their worst.
- 2 Words like 'ruthless' and 'cranberry' are holistic, but although they are partly meaningless in modern English, they are also both partly meaningful and perfectly meaningful as wholes: 'ruthless' operates like 'hopeless', even though 'ruth' does not operate like 'hope', and a 'cranberry' is obviously a berry. The whole set of berry words is odd: strawberry, raspberry, loganberry, elderberry etc. None of the first elements in the compounds is a word in its own right in any sense relevant to berries; the purpose of each first element is simply to point up a difference among berries. There are many examples of such half-analysable holisms, suffixed^{al} as in 'reckless', prefixal as in 'uncouth' or compounded as in 'cranberry'. People occasionally play with them, as when they suggest that someone

is feeling 'gruntled' rather than disgruntled, or looks 'couth' rather than uncouth.

- 3 This leads us on to the peculiarities of pronunciation and stress in certain kinds of holism. Take, for example:

receive,	reception,	receptive,	receipt
deceive,	deception,	deceptive,	deceit
conceive,	conception,	*conceptive	? conceit
perceive	perception,	perceptive,	*perceit

This set is odd in many ways, but the process of analogy by which it is formed shows that although meanings may be diverse, educated people would have little difficulty appreciating the principle involved in creating the set, and would have no trouble in saying the two non-words created by the paradigm. We could in etymological terms argue about whether or not the root 'ceive' has the meaning 'take' (via Old French from Latin capere 'to take'), but in synchronic terms 'ceive' has little if any meaning for most people. It still has a formative value, however. Among other things, it relates to another root form 'cept' and a third possible root form 'cei(p)t', and people appreciate the link. It also takes the accent, as do 'cept' and 'cei(p)t' in all these words. It may have little semantic meaning, but it has, as it were, phonaesthetic meaning. Similarly with the roots in all other two-syllable Latin verbs such as detain, explain, survive and so on. Here we have an example of roots still having work to do long after they have ceased to be formational bases. Comparably, academic people can make a fair guess at pronouncing gonadotropic or dioxyribonucleic, even if they do not know what they mean. They are acquainted with the phonology, if not the semantics, of such words.

- 4 We also have the problem of whether and how to analyse certain words. An encounter with the words 'slaver' or 'batter' in isolation condemns the observer to confusion. Such words are either holisms of the 'butterfly' type or they are fully analysable, but without context we can do nothing with them. If, however, we learn that saliva is involved and not ships, we can decide that slaver is a holism; if we learn that baseball is involved, we can see that batter is not holistic, although in cooking situations it would be. If we reversed the technique and introduced slaver only in the context of slavery and batter only in the context of cooking, then the relationships with mouth-watering and baseball would probably not occur to us at all. In such cases we cannot handle the words without context, but in other instances such as gather we would never propose an analysable 'one who gaths' or murder as 'one who murds', and so we do not even think of treating gather and murder as holistic. They are not. They are simply disyllabic, but their phonological nature is analogous to other words, some of which are holisms while others are analysable complexes, and we accentuate them all in the same way. On this same principle one could guess that the word womanize was created on a phonaesthetic analogy with humanize, and not on the semantic paraphrasing 'humanize : to make human' and so 'womanize : * to make woman' (except as a joke).
- 5 We also have the intriguing question of whether, ultimately, all complex words are holistic in some degree. Perhaps, when initially coined, they are simply the sum of their parts, but use in context soon demands more than just the adding of two or three constituents together. Suppose I invent a word of Greek provenance : *phonolatry.

An informed person can suppose that it means 'the worship of sound', and this is useful knowledge, but why and when and where is sound worshipped, and in what way? If I supply such information, then that becomes part of the meaning of phonolatry, and the word is now more than the sum of its parts. The same situation applies throughout the vocabulary, so that, for example, backstreet can be seen as meaning 'not a main street', but in the collocation 'backstreet abortionist' all sorts of additional nuances come into play.

- 6 Finally there is the possibility that, given an unique educational experience, set of inclinations and particular needs and aims at particular times, each user of a language is in a position such that, more or less deliberately, he or she can approach a complex word from an analytic or a holistic point of view. Thus, if one is dealing with a word like classify in terms of purify, simplify, centralize, harden etc, then one analyses it appropriately as a causative verb with a certain internal structure; but if one is dealing with it in a set which also contains arrange, order, class, codify and organize (three of which are non-complex words) then one may treat it as a solid like arrange, order or class.

Instead of regarding analysability or transparency as being one state in which complex words occur, and holism or opacity as being another distinct state, one is drawn to conclude that these are polarities, and that users of the language can choose (or be helped to choose) which pole they want to be nearer to on a given occasion.

This typology of holism as it stands clearly needs three tools of explication to help it work: paraphrase, gloss and contextualization.

Such a set of distinguishing procedures must be provided before one can confidently set about deciding the appropriate analysis for any complex word-form or word-group. The series of workbooks for foreign learners of English (McArthur, 1972 onwards) was created on the assumption that word formation and holism could be explained to learners by means of paraphrase, gloss and contextualization.

One footnote to this phenomenon is worth mentioning. We have seen at least one area with regard to which it is common: the vestigial prefixes of Latin-in-English. It can happen in such cases that a grammarian himself can contribute to the phenomenon. In the 1890s there was some argument as to whether contemporary or co-temporary was to be the accepted form of an adjective meaning 'occurring at the same time as something else'. These words did not just appear on paper: they had their individual pronunciations like today, and con- would have its weakened vowel in most forms of standard spoken English, while co- would have a full vowel and an accent. Henry Sweet threw his weight on the side of contemporary, because it was nearer the Latin (1891), contributing to the resulting comment of the Random House Dictionary in 196_Λ that co-temporary is 'archaic'. And yet the general evidence suggests that co-temporary is formed on a productive analogy with co-exist, co-partnership etc. Logically, Sweet should also have urged *conexist and *compartnership.

5.6 Derivational Paradigms

Sapir's word formulas deal with formation and Ross's with what has been formed. Neither adequately handles stages in formation or in stripping outer elements away from inner elements. Neither adequately covers the combinations of suffix addition, suffix exchange, accentuation alone,

accentuation together with affixation etc in the intricacies of English word-creation. Each says something useful about the principles involved, but in nowhere near sufficient detail. A formula therefore can only carry us so far, and what we must look for next in word-formation is a device for controlled construction and interpretation.

The device traditionally offered for this kind of thing is the paradigm, but it has been offered almost wholly for inflectional morphology, not for derivational morphology. Robins (1959), however, has considered the possibility of using the paradigm in order to depict derivation, pleading that the internal stability, the uninterruptability and the limited and regular extensibility of lexical words lend themselves to such an approach.

Traditional paradigms show how a base from an open-ended set of such bases relates to a closed set of inflections. 'Paradigms thus represent interlocking systems of grammatical oppositions within particular syntactic fields' (Robins, 124). The amo-amas type of paradigm is familiar enough and must follow its logical progression through to amamus and amant; derivational paradigms would be inherently different, however, because they would deal not in actualities but in potentialities - their full range need never be exploited. Such a paradigm will show what can be done, but will not attest what actually has been done. At the same time, however, the construction of such paradigms will depend upon a judicious interpretation of what has been done, what has occurred, and how we paraphrase or gloss the resulting complexes. Comparative research into potential paradigms of suffixation, for example, is made possible through the existence of modern reverse word lists.

We can illustrate this point with reference to Newman (1935). In his study of suffixation, he points to gaps in the even distribution of derivatives, thus:

normal	normality	normalize
modal	modality	_____
vocal	_____	vocalize

Newman assumed that these gaps were true for English at that time. How he attested this he does not say. In 1966, Webster's Third fills one gap by citing vocality and for the present writer in 1975 modalize seems perfectly acceptable. The word bound, for example, is 'modalized' in the sentence 'She's bound to go soon', meaning more or less 'She must go soon', where must is a modal verb. Such fortuitous gaps cannot therefore be regarded as absolute. They are not even interesting in terms of a modal of suffixation, because if derivational paradigms are worth postulating they stand as it were beyond time. The gaps are interesting, however, as regards how and when and where and why we choose to invoke the paradigm.

Entwhistle came close to such a view (1953) when he suggested that languages hover between completeness and incompleteness in their patterning, so that at any time one suffix, for example, may be dead while others are active, and the active suffixes 'have the effect of placing every word in a family which is partly realized and partly in posse'. Dictionaries in consequence list words with reputable authority behind them, but cannot reach all the words in use or likely to come into use. Lexicographers may hesitate to include all the conceivable products of a guessable paradigm without attestation, or may act on intuition rather than citation (cf Malkiel, 1968:263f), but inclusion

or non-inclusion in a dictionary bears little relation to whether or not a word will be invoked. Innovation occurs, and 'such successful innovation calls into existence words which are in posse and therefore in agreement with the genius of the language... Those which exist in posse can be called into being to meet a temporary emergency' (Entwhistle 76, 235 and 250).

This invocation or calling into being is what Bolinger calls a word's 'moment in history' (1968:53), and whether that moment is a flash or the start of a long and varied journey across centuries is an etymologist's concern, but irrelevant in terms of the paradigms. In other words, whether we talk about a nonce-word or an established word has nothing to do with principles of word-formation. This is a factor which has not been adequately considered by linguists.

In 1971 some preliminary work on programming derivational patterns into a computer was undertaken at the University of Edinburgh (Robert Boyer, Department of Artificial Intelligence, and Tom McArthur). The computer was asked to state the whole family of words formable on a given base, or any one member of such a family. This initial work suggested that an automaton for forming words was of some theoretical interest. Firstly, its products provided a challenge to the intuitions of the native user of the language as to the rightness or otherwise of a formation. Secondly, this intuition could be checked against a dictionary like the NED or with other native users. Thirdly, the need to program the computer accurately demanded great care in analysing what was at work in such paradigms. Thus, creating a set of members of a base type (such as all the bases entitled to the paradigm -al, alize, alization) necessitated not only first-order rules of a morphological

nature, they also demanded second-order spelling rules for properly combining the elements of the derived word. This system worked tolerably well for a number of paradigms of high productivity in the language, such as that given above, and its program is available with various printouts for examination and development.

Work with the computer and in creating vocabulary workbooks for foreign learners of English suggests that productivity in derivational paradigms is generally uneven, as Entwhistle proposed, but often much more regular than one might expect. A comparison with the NED and such historically-ordered word lists as The Chronological English Dictionary (Finkenstaedt, Leisi and Wolff, 1970) suggests that paradigms, like individual words, have their moments in time; that is, at different times different paradigms are fashionable, and for different disciplines and activities different paradigms are necessary. Thus, forms in medicine such as -otomy, and -ectomy do not occur elsewhere, while the form -ish extends to -ishness for colour adjectives (reddishness), but not for numbers when giving approximate times (sevenish but never *sevenishness).

Potentiality remains the most arresting feature of derivational paradigms. As Finkenstaedt puts it (1973:158): 'As sentences that have never been heard before are understood immediately because they are possible sentences of a specific language, these words are understood because they can be interpreted according to an internalized system. They are realizations of word-formation potentialities'. Thus, if we propose the paradigm -al, alize, alization, for certain Latinate noun bases in English we have two factors to consider:

formation on free word bases, and formation on bases which are not free words -

1 free word bases

centre: central: centralize: centralization

form: formal: formalize: formalization

2 bases which are not free words

re-: real: realize: realization

leg-: legal: legalize: legalization

Region, like centre, is a word which can be fed into such a paradigm, with the output regionalization. For many people several decades ago, this word would have been an unlovely and unnecessary addition to the language. It has, however, come into regular use in Scotland in the 1970s as part of a re-organization of local government. Let us suppose, however, that in another part of the world not region but area was an administrative term. The adjective areal is attested in Webster's Third, but no further along the paradigm. We might, however, see a bureaucratic need for *arealize and *arealization. It might be argued cogently that misanalysis as a-realize might militate against such a formation, and prevent its coinage. That is possible, but we cannot categorically rule it out for all time. Tomorrow's English cannot be legislated against.

Let us suppose that such paradigms as are suggested here and are more fully described elsewhere (Chapter Six and also in McArthur, 1972 et seq) become acceptable devices for use in discussing word-derivation in English. We cannot, however, forget the principle of holism, which occurs alongside all such activities. A derived word when

newly minted may well be no more than the sum of its parts, but if it exists for any length of time it will acquire further nuances of meaning unique to itself. Regionalization in Scotland, for example, has very special meanings which it does not have anywhere else in the English-speaking world, because of the administrative structure of the country at present. In addition to its standard paraphrase 'the process of forming (something) into regions', it must have the additional gloss 'as part of a two-tier level of local government in Scotland'. It is in glosses like these that many processes of holism probably have their beginning. If this is doubted, we can ask an Indian from Bombay what was meant in the mid-1960s by Maharashtrianization, a word which at that time was seldom out of the local English-language newspapers, but which stuns the user of English elsewhere when he meets it for the first time.

5.7 The Compounding of Words

A word-and-paradigm procedure can be used for describing and teaching important areas in English lexis, much as it has been used in the past to teach Latin and other languages, and much as Peter Matthews would like to see it used in discussing the morphology and formational principles of languages like Latin (Matthews, 1965a,b). The special link that paradigms have with the Classical language makes them attractive for handling Neo-Latin word-forms in English, but there is no inherent reason why they cannot be extended to cover material of Vernacular Germanic provenance. The principle behind cloud: cloudy: cloudiness and mist: misty: mistiness is no different from that behind form: formal: formality or even, in uniquely English circumstances, behind the set nose: nosy: nosiness as opposed to nose: nasal: nasality.

Compounding is a different matter, however, in that in its Germanic style it does not by and large lend itself to such patently neat patterns of extensibility. Additionally, compounding as a lexical and syntactic process has been obscured by some of the quirks of the English orthographic system, so that a compound such as teapot (described by lexicographers as a 'solid compound' because of the way it is written) is orthographically distinct from the compound coffee jug. No such distinction can or should be made in phonological terms as regards the accentuation (or stressing) of the compounds, or in semantic terms as regards their paraphrasable meanings. Phonologically, the two expressions have identical stress patterns and syntactically they can both be paraphrased in the formula 'an X for Y', so that we have 'a pot for tea' and 'a jug for coffee' respectively. Considerable work has now been done on the syntactic and semantic analysis of compounds in English (cf Lees, 1960; Marchand, 1969; the Gleitmans, 1970; Quirk et al 1973) and some work has been done on its use in language teaching (cf McArthur, 1972c; Marbe and McArthur, 1975).

The typological distinction offered in section 5.2 has a particular value in discussing compounds. Firstly, it enables us to recognise the purely orthographic problems of compounding for what they are: conventions which have grown up over the years rather than indicators of syntactic or semantic relationships between the elements constituting the compounds. Thus, we can see that a relationship exists between blackberry pie and black current jam, where adjectives have a special importance, and that there is another relationship between strawberry jam and orange marmalade, and that orthography is unimportant in all of them. Instead, phonological as well as syntactic and semantic considerations apply here, and the formational factors involved are of

quite a different order from derivational paradigms.

These formational factors are nevertheless highly regular. Given certain patterns or formulas such as the X of a Y, we can create, with the right word elements, innumerable generic expressions such as the skin of a goat = a goatskin, the top of a mountain = a mountain top, and with the minor variation of the X of Y we can get the erosion of soil = soil erosion and the removal of furniture = furniture removal. These formulas operate regardless of the structural content of the elements manipulated. Thus, goatskin is made of two free bases, goat and skin (which happen also to be monosyllabic), but furniture removal is made up of an element furniture which is a special kind of holistic noun relating to one meaning of furnish, and removal, which is a noun derived from remove.

That compounding is very different from derivation has been accepted for generations, but precisely why has not always been clear. Part of the reason for this is that, although different, the two processes have many common features, and tend to flow into each other. Like many other phenomena, the dividing line between these two is fuzzy. Consider the following:

- 1 We can recall for example Bloomfield's old maidish, which is based on a compound old maid, and not on maid alone: *maidish is not a standard word.
- 2 Holism affects both processes. Receive and butterfly are both holisms, though the first was created derivationally and the second by compounding.
- 3 Special compounding elements (or 'combining forms' as lexicographers often call them) may show features often associated more with

affixes than compounds. Hopeful can mean 'full of hope' and breakable can mean 'able to be broken', and on this paraphrasing basis they would appear to be elements in compounds, but their equivalents in hopeless (never 'less hope') and visible (needing translation as 'able to be seen') complicate matters and suggest a derivational interpretation. Worse, the presence of -ful in words like bashful and -less in words like ruthless and reckless make the whole thing a bad joke unless some such device as holism exists to account for them.

Two cardinal factors appear to operate in the process of compounding in English. They do not, however, necessarily occur together, although it is a satisfying fact that they often do co-exist. These factors are:

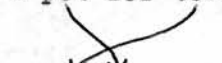
- 1 special stress, usually identifiable as the tonic placement

(cf Halliday, 1970; Marbe and McArthur, 1975), on the accentable syllable of the first element in a two-element compound, such as streetcar, blackbird, household, car factory and so on. This feature is not represented in the orthography but is for such compounds obligatory in speech. Many foreigners find it difficult to acquire and to interpret such special stress patterns.

- 2 a principle of a syntactic kind, conceived by some Chomskyans as transformational. This is the principle of inversion (McArthur, 1972:5 and 33), illustrable as:


a pot for tea

a teapot



a teapot made of china

a china teapot



Certain important uses of grammatical words (articles, prepositions etc) occur in paraphrasing such compounds, but it should be noted that these are not - to use the Chomskyan phrase - 'transformationally recoverable' (cf the Gleitmans, 1970). The ability to make such paraphrases was however established quite clearly by the Gleitmans as part of the 'competence' of the native user of English, even though the step-by-step creation of the compounds from a 'deep' paraphrase cannot always be traced ('recovered', in Chomskyan terms).

There is one feature which derivatives and compounds share which justifies their traditional close association, and that is their internal structuring, unless of course they are holistic. Particularly in careful speech, through stress and intonation, the native user can detect an organization within these composites which he can explain by means of paraphrases, much as follows:

1 derivation

recodification = (re(((code)ify)cation))
 = the process of codifying something again

2 compounding

a Savoy Hotel car park attendant
 = (a ((Savoy Hotel)((car park) attendant)))
 = an attendant in the car park of the Savoy Hotel

and also

a Savoy hotel car park attendant
 = (a (Savoy) (((hotel) (car park) attendant)))
 = an attendant in the car park of a hotel in Savoy

The complexity of compounded expressions in English, especially in newspapers, magazines and technical journals, can come as a surprise to

many when it is first discussed with them, and is a source of difficulty for foreign learners of the language. It is the same principle of inner organization as animates derivational structures, but differs in that it brings lexical words together in groups as novel compact formations governed by special stress patterns and explicable only by special paraphrasing techniques.

Compounding is also distinct from derivation in that it is, by and large, much more concerned with the creation of names; it is a generic labelling activity. As the Gleitmans observe (87): 'A compound, by being a compound, implies a name, some unitary character to the relationship among the elements'. I have chosen to follow the Gleitmans in keeping the non-specific term 'element' for each item which goes to the making of a compound, largely because such elements may on occasion be base words, but may also be derived words and indeed compounds in themselves.

There is one area, however, in which compounding and derivation are closely related, and that is in the less immediately obvious nature of material of Greek provenance in English. Greek compounds have their special stress patterns and a quality of internal cohesion that Germanic compounds lack. They usually operate, however, on a comparable principle of inversion, thus:

the study of life = $\begin{array}{ccc} \text{logy} & \text{bi} & \text{~~bi~~} \\ & \swarrow \quad \searrow & \\ \text{bi} + \text{o} & + & \text{logy} \end{array}$

the noun biology, however, can be turned into another noun biologist by losing -y and adding ist (without a stress shift), or into

an adjective biological by losing -y and adding -ical (with a stress shift). This special Greek quality of being able to add derivationally to a compound base often deceives us into forgetting that such bases are indeed compounds, and not collections of peculiar prefixes and suffixes. They are essentially very close to such formation as old-maidish.

In diachronic terms, a compound, whatever its degree of complex internal ordering, whether of Vernacular or Latin or Greek provenance, may be short-lived or long-lived. In newspapers, for example, headline compounds as in EGG THEFT DENIED BY VILLAGE YOUTH have a fleeting existence, whereas in technical language compounds such as 'internal combustion engine' take on a permanent lexical quality, and add to the meaning of the elements contributing to them. As with derivation, the success or failure of a compound in establishing itself in the language has no bearing on the principles of formation, but is of great interest in studying processes of holism and the development of lexical words through time.

5.8 Language, Lexis and Conceptual Universes

So far in this chapter the need to develop a typology of the English word and some means of discussing analysability and holism have held us back from the important basic question of semantics. Having proposed these various aids to the identification of 'words', it is necessary now to move on to a discussion of the vexed question of meaning. We can begin, however, by looking away from English-language linguistics towards a separate discipline: anthropology.

During the 1960s, the cognitive anthropology movement crystallized in the United States, in the work of Stephen Tyler, Harold Conklin, Charles Frake and many others. It has been concerned primarily with mapping the worldviews of different cultures, and belongs to a tradition shared by Edward Sapir, Benjamin Lee Whorf and Claude Lévi-Strauss. It is not possible here to analyse the movement in detail, but essentially it supports the 'weak' version of the hypothesis (often called the Sapir-Whorf hypothesis) that different cultures live to some extent in different conceptual universes, organizing reality in different ways. To illustrate this, we may take an example offered by Stephen A Tyler (1969:6), in a comparison between the culture of English-users and a South Indian aboriginal people. This comparison immediately comes up against the question of 'words':

'For example we distinguish between dew, fog, ice and snow, but the Koyas of South India do not. They call all these mancu. Even though they can perceive the differences among them if asked to do so, these differences are not significant to them. On the other hand, they recognize and name at least seven different kinds of bamboo, six more than I am accustomed to distinguish'.

The procedures adopted by these cognitive anthropologists have been given various names: semantic analysis, componential analysis and ethno-semantic. The first indicates a general interest in problems of meaning, of how people organize their world and talk usefully about it, and possibly why societies often fail to understand each other properly. Componential analysis suggests that they are interested in actual components or features of meaning, perhaps to set up a universal inventory of such components, common to all mankind. This is reminiscent of the aspirations in earlier times of Wilkins and Roget. Ethnosemantics

re-emphasizes the interest in peoples and cultures, rather than an interest in language for its own sake. Language is seen as both a means of getting at world-views, and an essential part of the cultures examined. Cognitive anthropologists ask questions, but are more interested in establishing the kind of questions that a society has asked itself than in simply getting answers to questions that our own world-view conditions us to ask.

The anthropologists quickly found that in using language as a gateway to separate universes they had a defective tool: the bilingual dictionary. Ethnographers, Charles O Frake points out (1969:28ff), have always spent time collecting labels for objects at which they have pointed inquiringly. This kind of activity results in monograph statements such as: 'Among the grasses (sigbet) whose grains (bunga nen) are used for beads (bitekel) none is more prized than Job's tears (glias)'. He considers that the unmotivated reader might justifiably mutter, 'So what?' But label-for-label the collecting continues till a bilingual register is achieved, and something appears to have been added to the sum-total of human knowledge. Harold C Conklin denies this, however (1969:41ff): 'While extant dictionaries and vocabularies do provide glosses and definitional material, many of the nontrivial, and often essential, semantic and contextual relationships obtaining among lexical items are often neglected or handled in an imprecise and unsystematic manner'. This is as true for the English side as for the side of the special language, in Conklin's case Hanunóo, a language of the Philippines. He points out that if a dictionary lists a Hanunóo plant name simply as '(distinct) kind of plant', the syntax is covered, but no semantic analysis is possible. 'Had I not modified this procedure, I would have ended up with more than 2,000 lexical items (including several hundred

referential synonyms) each labeled identically. While employing glosses like "tea" and "tobacco" ... proved useful in labelling familiar objects, the majority of those culturally significant Hanunóo designations referred to entities which to me were quite unfamiliar... For the ethnographer, the semantic structure of such folk classification is of paramount importance. Upon his analysis of it depends the accuracy of many crucial statements about the culture being described'.

The problem exists in reverse. To a Hanunóo inquirer looking at English, standard dictionaries might not be entirely adequate in providing clearcut semantic relationships among vegetable, plant, flower, bloom, fruit and seed, or turnip, radish, tomato, lettuce and pomegranate. Conklin considers that 'problems of analysing and presenting such structures in a succinct fashion may be of interest even to lexicographers who work only in relatively familiar cultural surroundings'.

The cognitive anthropologists have discussed their problems in terms of lexemes, lexical sets and lexical domains. A lexeme is any elementary unit whose meaning cannot be deduced from internal structure comprising simpler parts (even where there is such a structure). It is equivalent to our 'lexical word' as a base or as a holism. A lexical set 'consists of all the semantically contrastive lexemes which in a given, culturally relevant context share exclusively one defining feature' (Conklin, 46f), and therefore has something in common with Palmer's 'constellations'. The whole semantic range of such a set is its lexical domain. Conklin concedes that the initial establishing of domain boundaries can be a serious problem, not simply within one language,

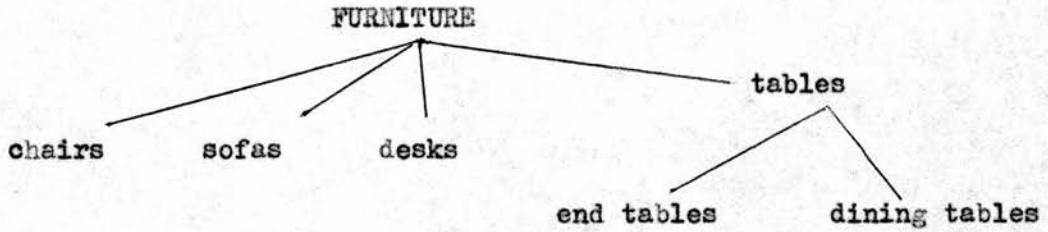
but also when seeking to correlate two languages and their respective domains.

Conklin and his colleagues have developed a set of logico-graphic devices which depart wholly from the principle of linear presentation. Conklin regrets that lexicographers have not explored the possibilities in such an approach, although he notes some tentative moves. 'Despite encouraging signs, I realize that most dictionaries will continue to be organized primarily as alphabetical indices' (56). The devices are based on an assumption that certain contrasts are fundamental in language and culture: the perception of similarities, differences, inclusion of one thing within another, exclusion of one thing from a group of others, possession of one or more common attribute among many things and so on. Tyler takes the English word chair as an example, the chunk of sound or lettering relating to objects with four legs, a seat and a back. No two of these are alike, but we manage to subsume them all under chair with reasonable success. Having done this, we also manage to label other things in similar ways, and catch them all under the head furniture. The devices serve to point up such fundamental relations. The three presentations which follow have been adapted from those Tyler gives (7ff), re-organized with English lexicography in mind:

1 The Taxonomy

A taxonomic presentation may be either tabular or branching, as in these alternative presentations of furniture. The same relationships of inclusion and contrast are shown in both devices. The taxonomy should not be taken as accurate or comprehensive for furniture, but simply as illustrative of the device.

FURNITURE				
chairs	sofas	desks	tables	
			end tables	dining tables



2 The Paradigm

The tabular arrangement of the paradigm has a venerable history in classical grammar, and requires a simultaneous reading of information along two axes, with 'lexemes' provided at suitable intersection points. Whereas the underlying structure of paradigms is perfectly symmetrical, lexical sets may not be structured symmetrically and so fortuitous gaps, overlaps and nebulous borders have to be allowed for. The paradigm which follows provides information about a superordinate word horse:

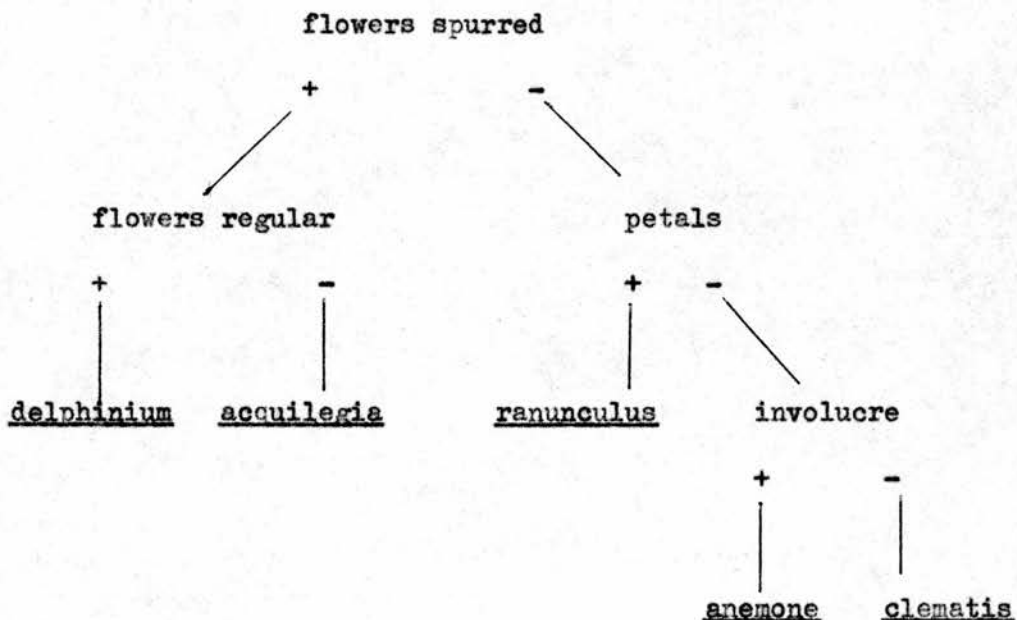
		<u>SEX</u>		
		<u>male</u>	<u>female</u>	<u>neuter</u>
<u>AGE</u>	<u>full-grown</u>	stallion	mare	gelding
	partly grown		filly	
	<u>immature</u>	colt		
	<u>new-born</u>	foal		

The markers SEX and AGE are each subdivided to provide special differentiations in terms of which lexemes can be registered. As it

stands, the paradigm does not indicate in what sense a gelding is neuter, because it models 'being' and not 'becoming'. From the linguistic point of view, the main problem is the status of the markers: are they to be taken as just other words of the language, or as abstract semantic features such as Roget attempted to perfect, or something of both, as Roget's turned out? In this paradigm they are differentiated from the lexemes which they register by means of distinct typefaces.

3 Binary Oppositions in a Tree Arrangement

This device provides paths along which the user can travel, accepting or rejecting a path according as it helps to show the presence or absence of the features in which he is interested. In this case, a tree arrangement (a hierarchy of antonyms) deals with kinds of flowers:



Thus, if a flower is not spurred, has no petals and no involucre, it is a clematis. Additional information may be required to describe the clematis more fully (with a picture perhaps), but cardinal points are

demonstrated, and again we can use typeface distinctions to separate the 'generated' plants from the mechanism which generates them by contrast.

Tyler concedes that these three devices do not exhaust the range of cognitive orderings, representing if anything only a small portion of cognition usable only where the number of properties involved are or can be restricted. He considers that in more complex domains, properties have a partial ordering (reminding us of Entwistle's views of patterning in language), characterized by discontinuous and partial combinations of various features, some binarily opposed, some intersecting, some included, some including and so on. He offers the devices, however, as a contribution to the as yet unstatable range of possible orderings.

Robbins Burling (1969:419ff) accepts that ethnosemantics is a salutary exercise in clearing the head of misconceptions about culture and language, but it has suffered in his view from a plethora of programmatic articles and a dearth of descriptions of whole systems or even definable subsystems. Among his criticisms, the most telling for lexicography is the problem of indeterminacy. He demonstrates his own problem of organizing 'plants' into one system in English, and considers that a multiplicity of orderings is possible. He presents the argument by outlining the following incomplete taxonomy, and then asking himself questions about it:

Plants							
Etc	Bushes	Flowers	Trees				
			Leafy Trees	Palm Trees	Needled Trees		
					spruce	hemlock	pine

He accepts it as reasonable that English-users distinguish trees from bushes and bushes from plants on the basis of size, then divide plants into flowers, vegetables and weeds. With regard to the category needled trees, however, he wonders how to place cedar, which is really neither leafy nor needled -- and should hemlock and spruce have a special subcategory 'short needled trees' to distinguish them from pine under 'long needled trees'. 'What is the essential "cognitive" difference between hemlock and spruce? Is it gross size, type of needle, form of bark or what? I do not know how to answer these questions, but they are the type of question that must be answered before any single semantic analysis can claim to represent the cognitive organization of the people, or even claim to be much more than an exercise of the analyst's imagination'.

Burling therefore counsels caution. There is an intrinsic difference between making useful analyses and penetrating to some God's-truth view of what goes on in people's minds, and the investigator is too willing to believe that he is capturing the psychological facts when all he is doing is 'tinkering with a rough set of operational devices' (427). Burling considers that ⁱⁿ (determinacy will always be lurking there, along with the possibility of other presentations, equally sound. But as long as the absolute is not substituted for the relative, no great harm can be done.

Dell Hymes (1969:428f) welcomes Burling's criticism as a valuable corrective to simplistic over-enthusiasm. Scepticism in his view should however be laced with humanity in handling any area where investigators are trying to chart values, orientations, attitudes, beliefs or any other notion which suggests something going on inside people. The

implication is that one cannot hold back simply from fear of claiming too much. The lexicologist can, however, stop with Burling while appreciating Hymes's point. He cannot assume the absolute determinability of meaning. Whatever the end-results of ethnosemantics, however, the lexicologist can thank the anthropologists for doubting the unidimensional item-by-item approach of the glossary tradition.

5.9 Structural Semantics

During the 1960s, a logical view of structural semantics was developed in Britain by John Lyons, at Cambridge and Edinburgh. The fullest exposition appears in his Introduction to Theoretical Linguistics (Cambridge, 1968:400ff). It is a logician's approach to semantic structure, a logician who discards the perennial philosophical question of whether there are such things as 'concepts' or 'ideas' behind words. Lyons also steps aside from the linguists' argument about whether there is a set of universal non-linguistic semantic features common to the organization of all languages. He notes that there may one day be justification for the assumption that such features exist, but 'such empirical evidence as there is available at the present time would tend to refute, rather than confirm, this hypothesis'.

The important factor to bear in mind, in Lyons's opinion, is that a semantic analysis can be made without entering upon a discussion of universal components or concepts, simply by treating the semantic components or features as other words of the language. If chair and stone can be said to denote 'physical objects', then physical object is itself an expression of English with definable relations. It need not be given, as in Roget (as well as in Katz, 1966), the status of a non-linguistic

conceptual or innate element. We can then say that semantics is concerned among other things with the relations that hold between chair and physical object, stone and physical object, and then between chair and stone. To get hierarchies we need not step outside language. It is of some interest to note here that Yorick Wilks, for practical purposes in considering text analysis by computer, came to the parallel conclusion that, except when people point to things, 'meaning is always other words' and to talk about 'the senses of words' is only a disguised statement of that fact (1972:86).

Lyons is sympathetic towards the American cognitive anthropologists such as Tyler. Referring to the analysis of kinship groups, colour, flora and fauna, weights and measures and other areas by the techniques outlined in the previous section, he notes that 'the results obtained have conclusively demonstrated the value of the structural approach to semantics, and have confirmed the pronouncements of such earlier scholars as von Humboldt, de Saussure and Sapir to the effect that the vocabularies of different languages (in certain fields at least) are non-isomorphic: that there are semantic distinctions made in one language which are not made in another; moreover, that particular fields may be categorized in a totally different way by different languages' (429). To illustrate this point he refers to Conklin's work with Hanunoo, to the way in which the speakers of that language distinguish colour on criteria of lightness, darkness, wetness, dryness, and not, like English-users, with reference to stock realia such as blood (for redness) and the sky (for blueness).

Lyons separates the grammatical from the lexical in 'words', and then passes to what he considers a basic dichotomy in semantics: the distinction between reference and sense. Reference has often been

regarded as a primary concern of semantics (as for example in Ogden and Richards, 1923). Lyons allows its importance for objects detectable through experience and for extrapolations such as unicorns and atoms which we agree to conceive of as existing (in some way or another). He points out, however, that for such words as good and intelligent referents are hard to find, and with other words like mountain and hill the referential boundaries are indeterminate. Reference only covers a relatively small number of words in any language, but sense covers all possible words. Relations of sense may be syntagmatic (sequential) or paradigmatic (contrastive) and they serve to place items in systemic positions uniquely their own. The vocabulary of a language contains a number of lexical systems whose semantic structure can be described in terms of sense relations alone, these relations holding between the ?\ lexical words themselves and not between their ~~senses~~. Lyons offers six such relations, refining in the process such traditional concepts as 'synonym', 'antonym' and 'inclusive term':

- 1 hyponymy The condition of superordination and subordination, so that tulip, daffodil and rose are (co-)hyponyms of flower. Characteristically, the extension of the superordinate flower is greater than that of any hyponym, while the subordinate (e.g. tulip) is intensive and possessed of a greater concentration of attributes than flower. Hyponyms contrast with each other and with their superordinate term ('hyperonym').
- 2 synonymy The condition of relative sameness, dependent on context, which can be viewed as symmetrical hyponymy, where for contextual reasons terms mutually imply one another, and tulip and flower do not contrast but are synonymous. Synonymy is characterized by the interchangeability of terms.

- 3 incompatibility The condition of contradictoriness, where, for example, red will contradict green in terms of a sentence frame like 'Mary was wearing a _____ hat'.
- 4 converseness The condition of opposites which reverse one another, as in rob/steal, buy/sell, husband/wife, in such formulaic propositions as 'X is the husband of Y, therefore Y is the wife of X'.
- 5 complementarity The condition of opposites which complete one another, as in the pair male/female. Characteristically, the denial of one implies in most contexts the assertion of the other.
- 6 antonymy The condition of opposites which can be graded or implies a scale, as in big/small in such formulations as 'If X is bigger than Y, then Y must be smaller than X'.

Sense relations of this kind are never absolute, but are context-governed. A structural organization aiming at presenting sets of words in a useful way would need to indicate the contexts for which such set relations hold true, and would need to take into account Lyons's notion of context, which includes:

- 1 the spatio-temporal situation in which speaker and listener find themselves
- 2 the actions they are performing at the time
- 3 any external events, actions, objects etc
- 4 the knowledge shared by the speaker and listener about what has been said earlier
- 5 the tacit acceptance by the speaker and listener of all the relevant conventions, beliefs, pre-suppositions in the speech-community.

Clearly, not all of this matters to the lexicologist, who must deal with the broader issues under 5, but may require to frame idealized or illustrative versions of 1 to 4 or look for citations. A distinction could

be made between immediate context (1 to 4) and background context (5 and perhaps 4). It is background context that the cognitive anthropologists are interested in, and lexicographers only look at immediate contexts (usually on paper) when they indicate a change in word use. One could say that lexical semantics, which deals with isolable words and not continuous speech or writing, is concerned with the effects that all contexts that have ever been have had on the words currently available in a language.

In his review of Lyons (Journal of Linguistics 9:1, Feb 1973), Haas emphasizes the logical foundations on which Lyons's structural semantics is raised. He praises the precision of the formulations, observing that if the outline does not yet present a structural semantics of language, it is at least a sketch of a semantics of logical language, or a true-or-false language of propositions reminiscent of Carnap in the 1930s. Hyponymy and the others are all defined by relations of implication among various propositions containing the lexical words concerned. They are restricted to truth-value relations, which Haas accepts as central, but necessarily limited in respect to natural language. Truth-value (as Mrs Piozzi pointed out in 1794) is not the only issue at stake in semantics, and the scheme proposed leaves no room for metaphor, jokes, idioms, slang, poetic language, paradox or levels of formality. Lyons avowedly takes a 'cognitive' as against an 'emotive' view of language, and although he briefly touches on distinctions such as 'colloquial' Anglo-Saxon and 'learned' Latinate, these are not developed. They have of course diachronic implications, and Haas notes that Lyons's position 'will be incapable of any easy connection with the all-obtrusive facts of continuous semantic change'. Logic stands apart from flux. Lyons states that semantic theory should indeed 'allow for' all the distinctions that

Haas mentions (462), but Haas argues that to allow for them is not enough: 'Semantic theory must account for them'.

Perhaps the most striking thing about the Lyons scheme, apart from its logical basis, is absence of time. It treats language as a state of fixed relations and implies that the semantics of English is a unity. Historical evidence suggests, however, that there is a semantic layering related to the multiple morphologies of English which have come together over the centuries. Lyons's category 'synonym' may serve to help us handle pairs like fatherly/paternal but cannot cope with father (a free word) and patern- (a bound base). Finkenstaedt (1973:64), in studying the influence of men like Sir Thomas Browne on English vocabulary, observes: 'Apparently the Elizabethans discovered the possibilities of etymological dissociation in language: amatory and love, audition and hearing, hearty welcome and cordial reception: these quasi-synonyms offer new opportunities for semantic differentiation. Two terms for the same denotatum: new connotations can arise, stylistic, poetic possibilities are offered when the new word is liberated from the restricted use in the language of science'. Lyons's system covers a great deal, but it does not cover this phenomenon.

5.10 Parallelism in English Lexis

At various points in this survey we have discussed areas or streams of lexical material in English which are of distinct provenance, principally Vernacular and Neo-Latin, and we have considered in historical terms how they came into a contrastive relationship.

English lexis is not a homogeneous mass, but rather a composite of

at least two masses, which appear ^{to} and behave differently but which share concepts. Thus, 'woman' belongs in one mass, while 'femin' belongs in another, both however referring to the same concept of a particular class of human being. There is nothing new in asserting this parallelism; various people at various times (as we have seen) have commented upon it, and even tried to make use of it. We have had the examples of the older lexicographers actively developing the parallelism; we have also seen how students of language such as Skeat, Sweet, Bloomfield, Crystal and Marchand have recognised and described the dichotomy. Additionally, writers such as Hogben (1964) have tried to use this idea to facilitate the teaching of Continental languages to the British and Americans. Hogben coined the phrase 'our hybrid heritage' for the phenomenon, and considered that it was an asset not adequately exploited. His purpose, however, was to show English-users that in their Vernacular language they had parallels with the Northern European tongues, while in their Neo-Latin language they had an entree into the Mediterranean tongues. As mentioned earlier also, Victor Grove vehemently argued that this hybrid asset has largely been left unexplained to the mass of English-users, so that many of us struggle less than successfully with Latin-in-English.

While conceding the historical phenomenon, and often, like Crystal, accepting the Vernacular-Latin dichotomy, linguists have on the whole shied away from a consideration of the impact of this heterogeneous circumstance on modern functioning English. We have noted this with regard to Lyons's structural semantics. Any attempt to account for levels of formality, for example, would, in my own view, require an assessment of Vernacular as against Neo-Latin and an appreciation of the nature of lexical bases, and would require a procedure of the kind being proposed here.

Hans Marchand (1960 and revised, 1969), in his diachronic and synchronic study of English word-formation, has made more effort than most to come to terms formally with the hybrid heritage, and discusses the lexicon (the total word-list of English) in relation to its members' 'foreignness' or 'nativeness', by which he means whether or not a word is Neo-Latin in type or Anglo-Saxon in type. In his study of formational principles he allows for a gradability of Neo-Latin material in English, as in this table:

1	<u>completely</u> <u>Neo-Latin</u>	spermarium, pre-retina, panopticon
2	<u>native in form</u> <u>but Neo-Latin in</u> <u>principle</u>	insecticide, spermaduct, pomiculture
3	<u>Neo-Latin in</u> <u>derivation, but</u> <u>analysable as English</u>	scientist, anelectric
4	<u>Neo-Latin in</u> <u>elements, but</u> <u>combined as English</u>	actional, hyper-sensitive

Marchand's gradation suggests a subtle shading out from Neo-Latin proper into something which Marchand accepts as English. He implies that what appears at first sight as a rigid dichotomy of words formed on a native basis or on a foreign or Neo-Latin basis is a procedural convenience rather than a genuine commitment to the ultimate separateness of languages. He notes, for example, that his foreign basis relates not just to a monolith called Neo-Latin, but to streams within Neo-Latin, which are Old Latin and Old Greek: 'Neo-Latin comprises Greek patterns as well and has frequently extended Old Greek patterns so that they are more rightly Neo-



THE UNIVERSITY *of* EDINBURGH

PAGE MISSING IN ORIGINAL

Latin than Old Greek'. A moment's reflection, however, makes it clear that Marchand is simply saying that Latin did in the past what English has been doing more recently: absorbing vocabulary from an external source. If a pattern can become more rightly Neo-Latin than Old Greek, then by the same reasoning a pattern can become more rightly Modern English than Neo-Latin. This is to all intents and purposes what has happened in cases 2, 3 and 4 of Marchand's graduated list.

What he has pointed out more formally than most is that a whole area of Old Greek vocabulary became Neo-Latin, and that in turn a whole area of Neo-Latin must have stood out in contrast against the 'native' Latin element, while in turn the new material in English continues to stand out in contrast against the older Vernacular material.

This is not an isolated phenomenon. The Dravidian languages of Southern India have for centuries been affected by an in-pouring of Sanskrit lexis; in the Middle East, Persian has been similarly affected by an in-pouring of Arabic lexis. Such parallel streams within these languages might be characterized simplistically as 'native' and 'foreign', but observers from a distance would naturally ask: How do the two co-exist? What role does each fulfil, if any? How in fact do they function in relation to each other? The same questions should be asked of English.

Lexical invasion has been studied with increasing formalism by the creolists (cf Hymes, 1971, in toto and 77ff), where the term 'relexification' has been used to describe the wholesale replacement of one lexis by another. This wholesale replacement relates specifically to the grammatical systems used by slaves transplanted to various places from Africa, but is not necessarily limited to such cases. There is evidence

to suggest, for example, that Marathi in India is a creolization of Sanskritic lexis onto a Dravidian grammatical system, with a wide spectrum of variation similar to the spectrum of variation found in Jamaican English.

X If we substitute the idea of a wholesale relexification with the idea of a partial invasion we have the cases of English, Persian and most of the Dravidian languages. In such cases the grammar largely stays 'native', along with one stream of vocabulary, used for general purposes, while a second stream develops, usually with special phonology and graphology, for religious, cultural and perhaps scientific purposes. Loosely, then, one might take the view that academic English is a creole of Neo-Latin. What does such a view entail?

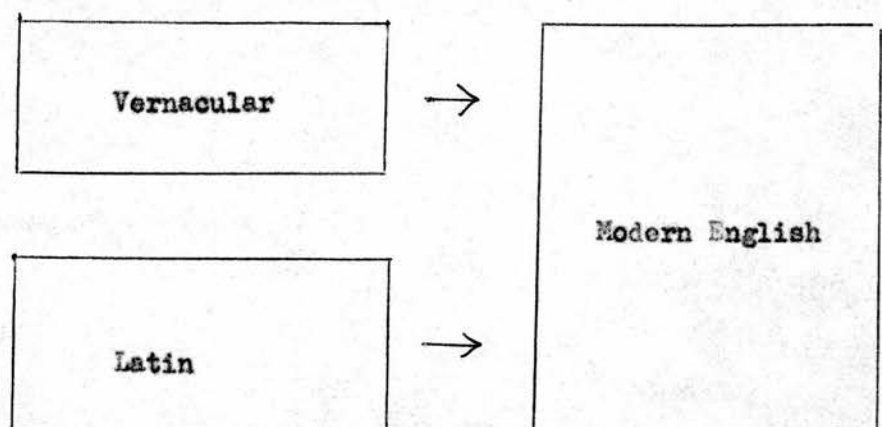
Recapitulating Sapir, we can establish the broad formational possibilities of English words as follows:

Vernacular	A + (b)
Latinate	A + (b)
	(A) + (b)

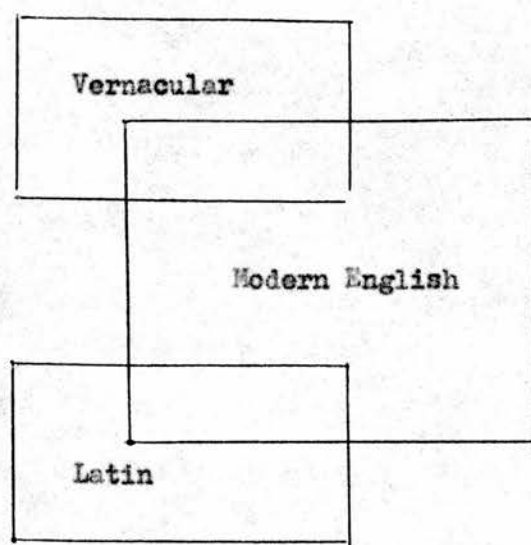
This view enables us to accept that the free word woman on the Vernacular level or tier of English is matched by a non-free base femin- on the Latinate tier, while the free Vernacular word shape is matched by the free Latinate word form. Therefore the factor which could bring such elements into alignment is not their status as free words, but rather their status simply as bases.

One of the probable reasons why dissystemic parallelism is not readily accepted is the traditional way of viewing languages and the

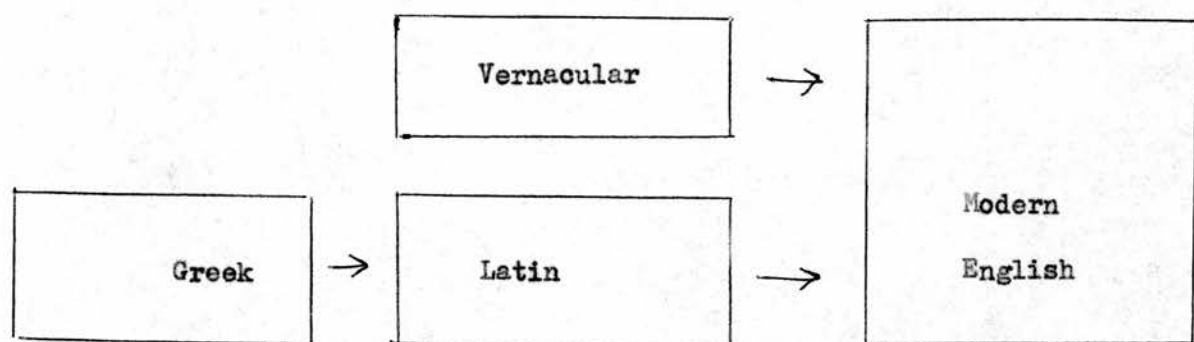
interaction of one language upon another. Languages are seen as co-existing but distinct entities, as follows:



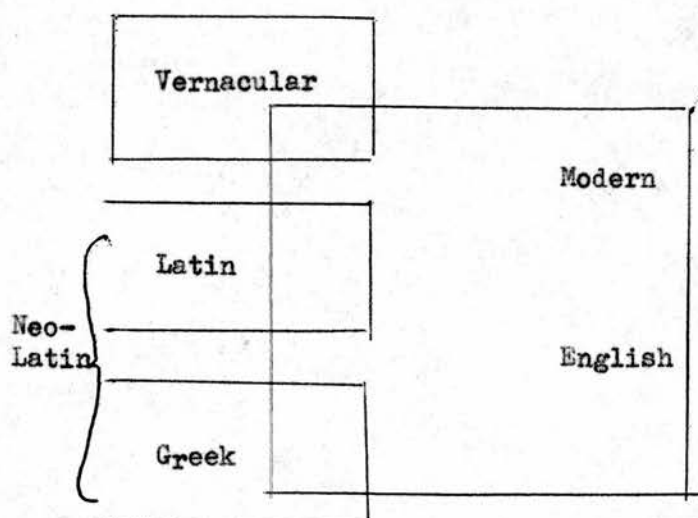
Here Vernacular and Latin are seen as distinct autonomous sources and Modern English is a distinct offspring. If, however, we propose that Vernacular and Latin can be taken as referring to something both inside and outside Modern English, then the model might be better shown as:



We have noted, however, that Neo-Latin underwent a similar kind of invasion from Greek at an earlier stage. We may say that a Greek stream entered Neo-Latin and via Neo-Latin passed to English. Using the model, this can be shown interestingly but inadequately as:



Inasmuch as Greek reaches Modern English through Latin it is also Latin. It would not surprise us to find, as Marchand makes clear, that Greek and Latin word-forming patterns have an intimacy which they do not share with the Vernacular. They have been together a lot longer. Suffixes of Classical provenance and patterning freely inter-relate, but neither inter-relates very freely (or with the same effects) with Vernacular material. At the same time, however, the Greek stream has its distinctive nature. It might be more proper to label 'Neo-Latin' as a totality, and to propose a three-tier approach to Modern English, thus:



Can any evidence be provided at this stage to suggest that this model has contrastive value, that it in any way reflects 'reality'? It is inherently paradigmatic, and with minimal adaptation can serve

to point up at least one distinction at work in the lexis of English, as follows:

V	over	SENSITIVE
L	super	
G	hyper	

where sensitive is a complex word-form (base -sens-) to which these three prefixes can be added. The result expresses a paradigmatic contrast between the prefixes in Modern English, not in any way relatable to the directional or locative uses of these elements in their languages of origin. The contrast is specific to Modern English. An attempt to paraphrase or gloss the new complexes helps to point to the functional differences inherent in the three tiers or streams of vocabulary:

<u>V</u>	<u>over-sensitive:</u>	too sensitive (in terms of some emotive statement)
<u>L</u>	<u>super-sensitive:</u>	very or supremely sensitive (in terms of some technical or cultural capacity, and approved of for that reason)
<u>G</u>	<u>hyper-sensitive:</u>	too sensitive, excessively sensitive (in terms of some medical or psychiatric statement, as part of a diagnosis)

Taking three bases alone, rather than prefixes, similar functional contrasts emerge:

V	woman	<u>as in</u>	woman-hater, womanly, womanish
L	femin-		feminine, feminism, effeminate
G	gynec-		gynecoid, gynecology

where, somewhat impressionistically but nonetheless significantly, we can say that the V is 'emotive', the L more 'cultural-academic', and the G more 'objective-precise-clinical'. All, however, relate to a single

concept which is best expressed as WOMAN, a lexical base (or 'lexeme', following Lyons, Lamb, Matthews and Tyler), or possibly what a Chomskyan might see as the underlying form of a super lexical formative with three quite distinct phonological realizations.

It is important at this stage to emphasize that any argument for the polysystemic nature of English lexis should be seen in functional terms rather than in purely etymological terms, although the two flow into each other. There is a danger that, because such terms as 'neo-Latin' have historical connotations, it will be thought that only strict etymological criteria can be employed in deciding what in modern English belongs in one area rather than another. We are attempting to get at popular usage, and people are notoriously not etymologists. The diachronic and synchronic dimensions of this theory are important, but only insofar as they are taken together.

Some examples may serve to clarify this point. Firstly, let us take the question of etymological purity. Most primers of language history tell the reader that such words as 'beef' are not Anglo-Saxon but are in fact of French provenance, of particular significance in Norman times. Of course this is so, but what is of interest today is the tendency for people to use words like 'beef' and other originally foreign words as though they were Anglo-Saxon. If a word-forming computer took 'beef' as a base for feeding into derivational paradigms and labelled it Latinate, it would seek to add any one of a number of Latinate suffixes to it, to form an adjective, and we would get:

beef: either *beefal or *beefic or *beefous

The language does not work that way. Instead, we need to fit these words into a paradigm that is of Anglo-Saxon (or better 'Vernacular') provenance:

beef: beefy: beefiness

These derivative words, however, have special meanings. For a child 'beefy' can mean 'like beef', as in 'a beefy taste', but for adults the primary interest is probably in a set along with fat, obese and so on. Similarly, nosy is different from nasal, as mouthy is distinct from oral and buccal.

We are discussing contrasts which exist in functional terms, and should take note of them just as we should take note of the opposite phenomenon, when words of distinctive Vernacular style are attached to Neo-Latin systems:

slenderize (to make slender)
kiddology (the 'science' of how to kid people)

These formations also have very special values, coined for commercial or for facetious purposes.

It is also of interest to note that words functioning on the Greek level need not necessarily be of Greek provenance, although they appear to need a similar morphology and a certain mystique. Sanskrit words, for example, and terms employed in certain sciences, become Neo-Latinized and in particular Hellenized:

1 Sanskrit terms

yoga	:	yogic	Sanskrit	:	Sanskritologist
karma	:	karmic	Bhagavat	:	Neo-Bhagavatism
Veda	:	Vedic	Dravida	:	Dravidian
Upanishads	:	Upanishadic	Arya	:	Aryanization

2 certain sciences

- Wyomingite (a mineral : geology)
- Finno-Ugric (a language family : philology)
- pasteurization (a process first undertaken by Louis Pasteur : medical science)
- macadamization (a process first undertaken by John McAdam : civil engineering)

One can at this point, recall the practice in traditional grammar-books for schools of listing 'Common Latin Roots', with brief translations in everyday English, with a few examples of how such roots appear in everyday words. Sometimes these would be 'Common Latin and Greek Roots', and they would be listed in alphabetical order, usually printed in italics. The theory offered here allows for a rather different treatment that is nonetheless sufficiently like the traditional practice to be recognisable and usable. To this end, I have developed a list of parallel bases ('triples' as they were called in the McArthur-Boyer computer program) intended to cover the commoner elements in the three areas that contribute so much to English lexis. Such a list is a convenient artifact, an abstract from the natural language and as such comparable in kind to Ogden's Basic English list. Informal experimentation with both native-users and foreign-learners of English over four years suggests that a list of such triples is taken by many people as a challenge as well as a reference tool. Once the principle is understood, it leads to morphological and semantic discussions which can encourage an interest in lexis and serves to co-ordinate the items of knowledge which many people have about the words they use. The following is taken from the E section of the list:

<u>Vernacular</u>	<u>Latin</u>	<u>Greek</u>
ear	<u>aur</u>	<u>ot</u>
earth	<u>terr</u>	<u>ge</u>
eat	<u>ed; vor</u>	<u>phag</u>
egg	<u>ov</u>	<u>o</u>
end	<u>fin; termin</u>	<u>tele; eschat</u>
eye	<u>ocul</u>	<u>ops, opt; ophthalm</u>

Use of the list with a dictionary or in discussion can raise and help settle questions about the relationship of ear-ache to otitis, aural to oral, terrain to geo-physics, eatable to edible to voracious, phagocyte and anthropophagy and even sarcophagus. Whatever its pedagogic side, however, it is possible that more formal tests can be made on people's attitudes to the 'low-, middle- and high-brow' words of English and further work can and should be done on the implications of this poly-systemic aspect of English lexis.

5.11 Models of English Lexis

The 19th century bequeathed to the 20th the idea of the complete dictionary list, and has produced such monumental attempts at comprehensiveness as Webster's Third International Dictionary which has earned from Uriel Weinreich (1964:30) the scathing comment that it is 'an anecdotal dictionary on a dinosauric scale'. In this he refers pejoratively to the endless line of thumbnail word-sketches that go to make up such a book.

Separately, Yakov Malkiel (1973:26) has warned linguists to be careful. Linguistic elegance may be possible in syntax, but it may never be possible in lexicology, 'because lexicology and lexicography are

inherently ponderous'.

This very ponderousness will lead in the not too far distant future to the word-archive as opposed to the wordbook. The ultimate dictionary is now being seen as a place rather than as a set of volumes (cf Sledd 1972:135; Twaddell 1973:218; Chapman 1973:310). The vision is enticing; Isidore and Johnson and all the other drudges down the centuries would probably have approved, but some questions have to be asked before the dinosauric activity is simply transferred lock, stock and barrel from page to databank.

There is one thing lexicographers of all kinds have agreed on: in compiling wordbooks, the apparent shapelessness has to be taken out of the lexicon. Of the two methods available for doing this - alphabetizing and the making of conceptual schemes -- the former has been greatly preferred, because it is easier. Its efficiency is remarkable, but at a price.

9/ The roman alphabet, like every other writing system, is an arbitrary achievement, concocted over long ages of tinkering. Its unmotivated nature is easily demonstrable, just by considering the different interpretation of the symbol J made by the users of English, French, Spanish and German respectively. One cannot express the alphabet as sound unless one knows the correlating code for a language, and for a language like English that correlating code is complex (cf Allbrow on English spelling: 1970). Consider in English alone the child's problem with 'g' in girl, gin, regime, tough, through and ghost.

If the alphabet were to vanish overnight in some one-sided cataclysm,

the lexis of English would be essentially undiminished, although dictionaries among other printed things would suffer. Such is the hold, however, of alphabetization for sound and for classification that it takes an effort to appreciate this point at the psychological level, where it matters. The ABC and words seem to belong together, although intellectually we can claim to see through the illusion.

There is nothing absolute in the connection between the 26 pigeon-holes and lexicography. Our current alphabet could just as easily have been (or become) 22 or 36 letters (or any other number) although the outcry against change would be great and we would feel uncomfortable with any loss or any addition. The widespread dubious response to Pitman's Initial Teaching Alphabet supports this statement. The roman alphabet, being arbitrary, could function just as well if arranged from Z to A or with S first; but more important than a given order is the fact of ordering. The alphabet as a classifier is uni-dimensional. It reduces everything which it classifies to a line and limits the possibility of clusters or groups. It imposes linearity where linearity is not natural.

The simplest consequence of this fact can be seen in such words as animal and zoo. They belong together, they occur together in the kind of constellation of words that Palmer discussed - and yet in every dictionary they appear an alphabet apart. Such a comment may seem laboured and trivial, but it is crucial to lexicography. The compiler, in putting animal under A and zoo under Z, is as it were repeating an ancient mantra: 'Lexis is a line, lexis is a line...' Some proof of the fundamentality of this mantra or diktat can be found in Zgusta's Manual of Lexicography, throughout which alphabetic ordering is an axiom,

the sine qua non of the craft.

The similarity between people's attitudes to wordbooks and sacred scriptures has already been touched on. It might not therefore be out of place to illustrate the alphabet diktat through a religious metaphor. In order to make a certain point to their followers, various prophets and apostles have said: 'God is our Father'. Psychoanalysts have discussed the relationship between a human father figure and the projection of a divine father figure, and so the social potency of the metaphor can be accepted. It depends on a straight Aristotelian proportion:

God : Mankind :: a father : his children

Such an analogy allows us to think certain things about the supreme deity, otherwise inaccessible; but at the same time, by accepting this metaphor and no other we impose certain restrictions on our conception of God. People of the Judaic-Christian tradition recoil at an outright 'God is our Mother', and wuld boggle at 'God is our Uncle'. Many Hindus however assert the second metaphor, and in some remote society the third may well have been invoked. In Europe, however, the sheer weight of tradition supports the Father-metaphor.

Much the same applies to lexis-as-line. The scholar like the prophet wants to catch his elusive subject. The writing system suggests alphabetization just as the family system suggested 'father' to the prophet. In accepting the alphabet the compiler achieves much, but pays the price of a similar restriction. He limits our and his capacity to see the lexis as anything other than a line, just as it is hard for God to be anything other than Our Father which art in Heaven. A theologian

talking about God-as-Father may know (intellectually) that he is dealing in metaphor for didactic purposes. The mass of adherents are not accustomed to subtleties, however, and psychologically for them God is their Father. Quite a few theologians may also settle for the device as having so much of 'truth' in it as to be effectively the truth. In the same way, a lexicographer studying the lexis may know that it isn't a line, but may also consider that there is so much to be gained from proceeding as if it were that he declines to question further. He may well aver that, for tidiness, it ought to be a line, and begin to think of his own list as a feeble reflection of some Platonic 'True List' always just beyond mortal grasp.

Lexicography can be compared to science as well as religion. In science development is less by smooth accumulation of 'facts' as by sudden upheavals of theory (cf Thomas Kuhn, The Structure of Scientific Revolution, 1962). One of the more spectacular revolutions has been the switch from Ptolemy's geocentric to Copernicus's heliocentric universe. There is a certain sense in which the Ptolemaic view of the universe is 'true'. From the individual standpoint there is nothing unreasonable in proposing that the earth is the centre of the universe and flat, that the sun revolves round it, that it is much larger than the sun and so on. Certain practical activities like navigation work very well on such assumptions, and so does the everyday human mind. Most educated people are Copernican in their intellectual commitment, but Ptolemaic in their day-to-day commitments. They behave as though the world were flat and central to the universe, although they know it isn't 'really'. We might say that intellectually they are Copernicans and psychologically they are Ptolemaics.

The analogy is useful for dictionaries. Because of the alphabet we are conditioned to think of the lexis/lexicon as a list stretching more or less without end, capable of insertions and expansions but little more. Having this convenient 'Ptolemaic' view, we can still see as Palmer saw that words belong in 'constellations', a term which suggests multi-dimensional groupings. We can detect structure of a multi-dimensional kind, but we cannot express it. We can only express linearity, because technologically we are not 'Copernicans'. The Copernican structured view of the lexicon is encouraged by the whole drift of linguistics since Ferdinand de Saussure (1916), but linguists discussing 'the lexicon' behave generally as though it were a list, a line of items to be slotted into the sentence-making machine. If this is so with linguists, it is also the case with ordinary laymen, who expect glossaries. Even scholars who appreciate structure finally produce glossaries.

The model of lexis-as-line has, however, been challenged. James Murray, third editor of the NED, challenged it as a way of looking at any language. The words in which he expressed the alternative are vividly metaphoric, and appear in the 'general explanations' to vol. I of the NED in 1888. I have divided the original continuous paragraph into eight new paragraphs in order to highlight the various points made:

'The Vocabulary of a widely-diffused and highly-cultivated living language is not a fixed quantity circumscribed by definite limits.

'That vast aggregate of words and phrases which constitutes the Vocabulary of English-speaking men presents, to the mind that endeavours to grasp it as a definite whole, the aspect of one of those nebulous masses familiar to the astronomer, in which a clear and unmistakable nucleus shades off on all sides, through zones

of decreasing brightness, to a dim marginal film that seems to end nowhere, but to lose itself imperceptibly in the surrounding darkness.

'In its constitution it may be compared to one of those natural groups of the zoologist or botanist, wherein typical species forming the characteristic nucleus of the order are linked on every side to other species, in which the typical character is less and less distinctly apparent, till it fades away in an outer fringe of aberrant forms, which merge imperceptibly in various surrounding orders, and whose own position is ambiguous and uncertain.

'For the convenience of classification, the naturalist may draw the line, which bounds a class or order, outside or inside of a particular form, but Nature has drawn it nowhere.

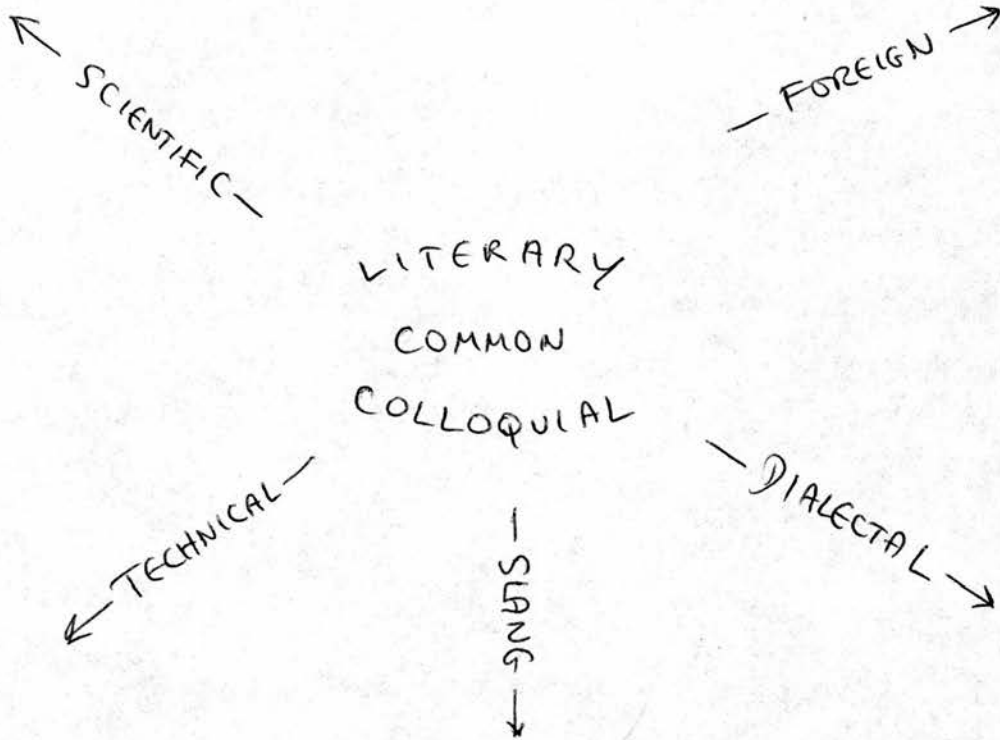
'So the English Vocabulary contains a nucleus or central mass of many thousand words whose "Anglicity" is unquestioned; some of them are only literary, some of them only colloquial, the great majority at once literary and colloquial - they are the Common Words of the language.

'But they are linked on every side with other words which are less and less entitled to this appellation, and which pertain ever more and more to the domain of local dialect, of the slang and cant of "sets" and classes, of the peculiar technicalities of trades and processes, of the scientific terminology common to all civilized nations, of the actual languages of other lands and peoples.

'And there is absolutely no dividing line in any direction: the circle of the English language has a well-defined centre but no discernible circumference.

'Yet practical utility has some bounds, and a Dictionary has definite limits: the lexicographer must, like the naturalist, "draw the line somewhere", in each diverging direction'.

Murray supplemented his remarks with a circular diagram:



He also suggested that the domain of common words widens out differently for different people, not in an even circle, but each in the direction of his own experience through reading, business, residence and the like.

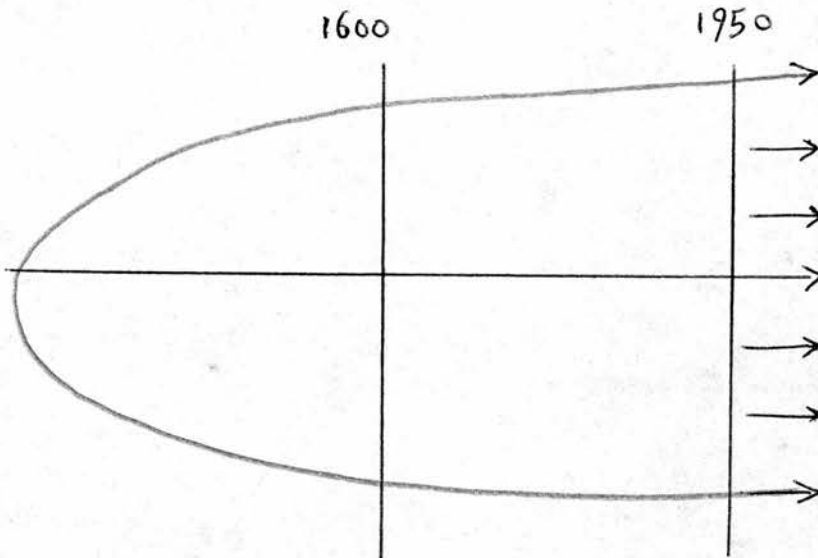
Finkenstaedt (et al, 1973:54) notes that 'there are not very many "models" available for the description of the system of words in general or the system of a particular language. The ingenious diagram devised by Sir James Murray for the OED ... is unsurpassed in its simple and yet comprehensive structure. It is an almost purely synchronic model ... The depiction of vocabulary as a closed circle (with the implicit

assumption that it is an expanding circle) is of limited value only. A clear terminological distinction should be drawn between a vocabulary... i.e. the lexical items of a defined corpus (a text, the words of a poet, a dictionary) and the lexicon ... i.e. the totality of words available to a given person or at a given time. The lexicon cannot of course be defined numerically like the vocabulary but statements about the quantitative relationship between a given vocabulary and the corresponding lexicon can be made'.

He considers that a statistical analysis by computer of the vocabulary in the Shorter Oxford English Dictionary indicates the relative success of the SOED in reflecting the lexicon of English accurately. He is correct in assessing Murray's diagram as synchronic or stati^c, but Murray did not intend it to be taken in this way. In print he adds the diachronic axis as follows: 'The living vocabulary is no more permanent in its constitution than definite in its extent. It is not today what it was a century ago, still less what it will be a century hence. Its constituent elements are in a state of slow but incessant dissolution and renovation'. Murray drew a handy diagram to show what the vocabulary of English was like: he did not propose to act on it, and proceeded to compile a list of mammoth proportions. Taken to its logical conclusions, Murray's diagram is much more than a Circle Model: it is a Galaxy Model, with word-particles in shifting swirls, individuals fading out, replaced by others caught by gravity or generated from the stuff of language, some particles in a slow or rapid journey from perimeter to centre, from centre to perimeter, the whole never still, always indistinct.

Finkenstaedt offers an alternative to the circle as a model (159f),

a parabolic growth model to take account of the time dimension and to catch the idea of potential words at every moment during the development. The following diagram is a simplification of the Parabola Model, a kind of sideways view of Murray's circle.



a) 'The speaker can be imagined on a point on the x-axis outside the parabola ... The size of the lexicon cannot be described in a finite number, but comparative calculations on the basis of vocabulary and potentiality should be possible for particular languages at a given time'. Points on the x axis may indicate either dates or guessed-at quantities of words. The time dimension is crucial in the Parabola Model because a language pre-exists and outlasts the individual reflecting on its constituents and potential. 'As a point of reference the lifetime of a speaker is too short,' and the individual is not very good at saying what he does or what he can do with the lexicon, in judging wellformedness in words, usage or appropriateness. 'Thus the lexicon has to be investigated with the help of a chronologically ordered vocabulary.' It is in this that Finkenstaedt and his colleagues make a major novel contribution, a rather different kind of linear dictionary: A Chronological English Dictionary (1970), computer-compiled with words

entered in terms of Bolinger's moment in history (as far as the evidence can be relied on).

Murray and Finkenstaedt have attempted to be Copernican in theory, even if neither of them could in practice escape the alphabetic list. Finkenstaedt assumes that we can largely reconstruct the lexicon over time. We cannot. We can only dig among the fossils of the past, and only in the written part of the language. Additionally, in the state of our knowledge at the moment of how the mind works and of how we handle 'words' in neuro-physiological terms, it is not yet possible to offer convincing models of the multi-dimensional inter-relationships of lexical material. Each individual mind has its own system, and we lack the information to help us generalize about what is common to the minds of all English-users everywhere.

One suspects, however, that little progress will be made towards a consensus model of the lexicon that has some psychological validity if we do not come to terms with the nebulosity implicit in both Murray's and Finkenstaedt's views of the lexicon. It reminds us of Entwistle's comments on the incompleteness of language patterns, of the doubts of the critics of cognitive anthropology, and appears in more recent linguistics with Lakoff's remarks about 'fuzzy sets' (1973:144): 'As any lexicographer knows, natural language concepts are fuzzy; the boundaries are not clearcut'. Absolute models are impossible; we can aim only at better or worse versions.

If that is so, then instead of uncritically continuing to make ABC lists, we should be considering the possibility, not so much of supplanting them, but of supplementing them. Alphabetic lists may be

the best device for certain purposes, while a circle or cluster model may be better for other purposes. Lexicography should widen its horizons to a range of options available to a compiler, and devices such as those suggested in this chapter have their place in the search for multi-dimensional representations of the relationships among words.

5.12 Conclusion

In half a millenium of lexicographic work the amount of labour put into the elucidation of English words has been immense. As a social and commercial programme the work still continues, and is comparable to any other major undertaking in the sciences, the arts or in politics. Of the educational side of lexicography, Raven McDavid has said (1973:5) that since the Renaissance at least 'the need to provide information about language for the uninformed and socially insecure underlies all subsequent lexicography'. His wording suggests a kind of scholarly paternalism, but the ambitions of such scholars have been surprisingly democratic, and can be seen as part of the process of universalizing literacy and knowledge.

Yet, as we have seen, the process has been remarkably unself-conscious. Uriel Weinreich says of it (1962:26): 'The indifference which lexicography displays towards its own methodology is astonishing. Perhaps lexicographers are complacent because their product "works". But it is legitimate to ask in what way it works except that dictionaries sell'. Thomas Finkenstaedt (1973:20) adds: 'The existence of the OED, Webster's Third International and other great dictionaries is usually taken so much for granted that not much thought seems to have been spent on investigations of their structure or their background: Dictionaries

are there to be used and not to be analysed. There are few substantial reviews of the "big" dictionaries and there are not very many studies of individual dictionaries or the history of dictionaries in general'.

The present study is a move towards an adequate reply to these cogent observations. It can serve as a stepping-stone towards a fuller study of lexicography, and also of lexicology, a somewhat overshadowed area in linguistics at large and in the linguistics of the English language in particular. Of the inter-relationship of these two activities Zgusta (1973:14) says that 'whereas lexicology concentrates more on general properties and features (of the lexicon) that can be viewed as systematic, lexicography typically has the ... individuality of each lexical unit in the focus of its interest'. This is an interesting proposition: an -ology for the generalities and an -ography for the minutiae, but only if it allows for a difference in kind between the two, as well as the possibility of overlap. Lexicography possesses elements which make it an art, while lexicology is a branch of linguistics and as such is a social science, part of whose job is surely to review lexicography.

Lexicographers are now faced with the possibility that the next great store of lexical data will be a computer complex and not a set of volumes on a shelf. We move towards the era of the Dictionary as a Place. They can therefore no longer afford the luxury of benign neglect towards the theoretical side of their craft. Otherwise, they could find themselves starting on their 'Archives' without knowing what it is they want to put on file, much as the word-counters started their work. Similarly, there is a good case for linguists to continue developing a theory of language which accepts within individual languages those

psychologically satisfying bits and pieces that everybody knows are 'words'.

In the latter stages of this survey an attempt has been made to create the typologies and techniques necessary to begin handling English words. Firstly, following Lyons, Lamb, Matthews and others, a basic disambiguating typology has been offered for the term 'word' itself, with regard to orthography, phonology, morphology, lexis, grammar and statistics. Next, a distinction has been made between morphology and syntax which, following Jespersen, avoids the pitfalls of assuming that there is a smaller-larger relationship between the two. Given that they are distinct views of the same phenomena, we can take a morphological 'external' view of the language in which an analysis of segments and processes can be undertaken. For such an analysis we need certain pre-requisites, and for these I have drawn heavily on the proposals of Vendryes in Europe and Sapir in North America, proposing a distinction between 'lexical bases' (the semantemes of Vendryes; the radical elements of Sapir) and 'process markers' operating upon them (the morphemes of Vendryes; the grammatical elements of Sapir). We can adopt the formulas of Sapir and of Ross the etymologist for handling concatenation and process in word-formation, and also develop the ideas of Entwistle and Robins in order to explore the possibilities of paradigms for the derivation of words. A full demonstration of the possibilities inherent in such an approach is given in Chapter Six. I have also outlined ways in which the ideas of Lees and the Gleitmans can be developed in order to analyse patterns for the compounding of words.

At this point we move from 'external' and morphological, to 'internal' and syntactic, to a consideration of why people create new word-formations,

and how they use words generally. Both derivation and compounding can be 'explained' syntactically and for general semantic purposes by means of paraphrases and glosses, expressing the special relationships among their elements, their bases and their process markers. Where the relationships are transparent, 'bases' can be discerned functioning in current English; where they are opaque, historical analysis and functional comparisons can reveal the existence of 'roots' and 'holisms', and these latter have their own typology. Holisms can be seen to have a relationship with the past of the language, with earlier times when they were still analysable. They demonstrate cogently the need in lexicological analysis to combine synchronic with diachronic views of the ^adate being studied, much as Marchand has done. At the same time, however, we have seen that holistic usage also co-occurs with analysable usage in the on-going language.

The provisional set of typologies and techniques offered here are, consequently, a synthesis and development of a number of past statements by a variety of linguistic scholars. It is an eclectic approach, belonging to no specific school or movement in linguistics, but is offered as an entrée into a problem - the nature of lexis - which is central to the study of natural language.

The approach demonstrates that the lexical aspect of language is extremely complex, without, however, retreating from it in despair. It also stresses the feats of creativity involved in the use of lexical bases and the markers which operate upon them. Its synchronic-diachronic element, a kind of binocular vision, allows us to combine knowledge of the past with analysis of the present - and may even allow certain cautious guesses at future lexical possibilities. It also makes possible a study of the functions and relationships of the various streams -

Vernacular, Latin and Greek - that dominate the lexis of the English language. Many have felt, like Grove and Hogben, that we have not paid adequate attention to the 'hybrid heritage' of English, and in this chapter I have proposed certain provisional descriptions of and solutions to this problem.

In no analysis of a language's lexis, however, can the significance of the conceptual universe inhabited by its speakers be left out. In making this statement it may be said that I have adopted a weak Sapir-Whorf position, claiming that language and our sense of reality - our worldview - are intimately interwoven. Cognitive anthropologists such as Tyler, together with the structural semantics of Lyons, have been introduced to indicate certain lines of investigation developing in this area, but also the wealth of work still to be done: in analysing the intricate systems of relationships obtaining among the lexical bases of any language, or, in particular, among the jostling 'words' of English. Murray's and Finkenstaedt's models, though still pre-theoretical, serve to emphasize the complexity of the task ahead.

It is proposed finally that past traditions provide an excellent foundation on which to build a future understanding of English lexis, its place within the language, and the place of lexis in language generally. Mutual co-operation among lexicographers and linguists in the next few decades will no doubt help set the stage for the next half-millennium of harmless drudgery.

6 Some Derivational Paradigms for Suffixation in English

Introduction

This chapter differs from the preceding chapters in that it is not a discussion or argument developed stage by stage, but rather presents a model of part of the English language. The model is an attempt not only to chart out an area which has not been systemically handled before, but also to demonstrate how certain principles and techniques proposed in chapter 5 can be used in clarifying relationships at work in a specific complex area of word formation and word use.

The principles are: that the words of English can be described in terms of 'lexical bases' and 'process markers' acting upon these bases, whether or not such bases are 'free' words; and that a root-and-base distinction and the concept of holism can be usefully applied to situations where some words appear less than amenable to systemic treatment, because of some kind of semantic opacity, than others which are similar to them. The techniques are the derivational paradigm, and a system of paraphrasing for the meanings of words that are formed by inserting bases in paradigms, and of glossing for related words which are now holistic but at some earlier stage in the language were probably formed via such paradigms.

Suffixation has been chosen because of its inherent complexity, but it is only one of a number of areas of lexical interest that might have been chosen for the purpose of demonstrating the principles and

techniques. Other such areas are prefixation, compounding and the internal patterning of phrasal verbs. Although these are not discussed here, reference may be made to McArthur (1972a,b,c), McArthur and Atkins (1974), and Marbe and McArthur (1975) for indications of how such other areas can be handled.

The model is still only an interim statement, but it is relatively detailed, though not exhaustive either in coverage or in the preciseness of its detail. Nevertheless, it is offered as a means of replacing the standard lists of 'suffixes' in such works as Marchand (1960/69) and Quirk and Greenbaum (1973) with something more expressive of the multiplicity of forms available in the language.

The layout and ordering of the material is arbitrary, but, it is hoped, provides a framework within which a range of statements can be made about lexical and morphological organization that might not otherwise be compactly made at all. The model consists of 15 paradigms which are described variously as:

- 1 of Vernacular or Neo-Latin provenance, developing the proposals in Chapter 5 on the di- or trisystemic nature of English lexis
- 2 stating whether a paradigm is based on nouns (denominal) on adjectives (deadjectival) or on verbs (deverbal), following Marchand's and Quirk's terminology.

Within each paradigm there are six divisions of information:

- 1 a statement of the type of base(s) which can exploit the paradigm
- 2 where necessary, a statement of the derivatives that can also wholly or partly exploit the paradigm
- 3 a diagrammatic presentation of the paradigm itself

- 4 listed examples of word-complexes created in the paradigm
- 5 a set of definitive paraphrases showing how complexes created in the paradigm can be interpreted
- 6 a set of notes which add qualifications to the general utility of the paradigm where necessary, or additional comments about sub-paradigms and anomalies, and which help to allow for the 'fuzziness' of natural language.

The model as it is presently arranged does not incorporate directly into the paradigms information on stress- or accent-shift in derivatives. Many points connected with this important matter are, however, added in the notes to the paradigms. It is an interesting point that suffixes of Vernacular provenance never cause stress- or accent-shift, whereas, although many suffixes of Neo-Latin do cause it, they do not all do so, and do not all operate in the same way. Information on this matter has been left out of the paradigms because they are already considered complex enough for presentation at this stage, but not because it is impossible to add that information. The table below is one method of listing the shifting of stress or accent. It does not, however, attempt to show the adaptation of vowel quality which usually accompanies such shifts, because this work is not primarily phonological. More information on vowel-quality adaptations is available in Marbe and McArthur (1975). In the list below the numbers of the paradigms in which the various suffixes can be found are given in brackets after each suffix or suffix-cluster, but it should be noted that most of the suffixes in the fourth column are not at present incorporated into the model, because of their relatively low level of productivity, and the same principle applies to the suffix -ate in the first column.

stress falling:

1 <u>two syllables</u> <u>earlier than</u> <u>the suffix</u>	2 <u>one syllable</u> <u>earlier than</u> <u>the suffix</u>	3 <u>on the first</u> <u>syllable of</u> <u>the suffix</u> <u>itself</u>	4 <u>on the only</u> <u>syllable of</u> <u>the suffix</u> <u>itself</u>
-al (5,7,9,10,12)	-an (6)	-acious (8)	-ee
-ar (5)	-ian (6,10)	-acity (8)	-eer (1)
-ate	-ial (5)	-ation (7,12,13)	-esque
-ine (8)	-eal (5)	-atory (12)	-ese
-ize, -ise (5,6,10,13)	-ual (5)	-ition (7)	-ess
-oid (11)	-ic (9,12,14,15)	-ution (7)	-ette
-ous (8)	-ical (9)	-atic (9)	
	-ics (10)	-etic (9)	
	-id (8)	-icious (8)	
	-ify (12)	-istic (14,15)	
	-ile (8)	-mental (5)	
	-ion (7)	-ental (5)	
	-ity (5,7,8,9)		

NOTE Where there is only one syllable before a suffix listed in the first column, stress falls on the preceding syllable: 'legal, 'solar, 'feline, 'cuboid etc, except for verbs in -ate, where the stress falls on the suffix itself: pla'cate, e'quate etc.

The source data used for the analysis leading to this set of paradigms was collated from:

- 1 The Normal and Reverse English Word List (1963), compiled under the direction of A F Brown, the University of Pennsylvania.
- 2 Jespersen, Onions, Marchand and Skeat, as listed in the bibliography.

- 3 three editions each of the Sunday Times, Time Magazine and The Scotsman in early 1970.
- 4 The Random House Dictionary (RHD) 1966 which lists derivatives extensively and clearly.

Abbreviations and symbols used are:

<u>n</u>	noun	+	add	
<u>a</u>	adjective		convert to	?
<u>v</u>	verb	=	is equivalent to	
T	tonic placement	...	and others of the same type	
			not attested in personal experience or in the RHD	

Paradigm 1 basic deverbal pattern (Anglo-Saxon provenance)BASE TYPES

- 1 Those simple regular verbs, mainly monosyllabic and Anglo-Saxon in provenance, for which the paraphrases work meaningfully:

plan, walk, dance, jump...

- 2 Comparable irregular verbs whose past-participial forms supplete the regular form in +ed:

teach, run, think, swim...

- 3 Comparable mainly disyllabic holistic verbs of Latin provenance:

receive, retain, respect, control...

collect, direct, inspect, conduct, act...

DERIVED TYPES

Forms such as:

- 1 sharpen, blacken, freshen, brighten... (Paradigm 2)
 2 classify, simplify, purify, rectify... (Paradigm 12)
 3 fertilize, centralize, sterilize, humanize... (Paradigm 13)
 4 dictate, regulate, incubate, percolate... (Paradigm 7)

PARADIGM

<u>v</u>	<u>n</u>	1 <u>+er</u> (agentive) (instrumental)
		2 <u>+ing</u> (gerundive) (actional)
	<u>a</u>	1 <u>+ing</u> (present-participial)
		2 <u>+ed</u> (past-participial)

EXAMPLES

plan	teach	receive	select
planner	teacher	receiver	selector
planning	teaching	receiving	selecting
planning	teaching	receiving	selecting
planned	(taught)	received	selected

DEFINITIVE PARAPHRASES1 +er, six subtypes:

- 1.1 an occupational agentive noun relating to simple tenses of a base verb:

intransitive verbs

She dances as an occupation. = She is a dancer.

transitive verbs (1)

(object noun retained in a compound formation, with tonic placement (T) on the first element of the compound)

He makes glass as an occupation. = He is a ^Tglass-maker.

transitive verbs (2)

(object noun deleted as understood, unless special contrast is required)

He drives cars/vehicles. = He is a driver.

He drives ^Tbuses, not ^Ttrucks. = He is a ^Tbus-driver, not a ^Ttruck-driver.

(NOTE the singularization of buses to bus and trucks to truck)

- 1.2 a habitual agentive noun relating to the simple tenses of a base verb, often with an intensifier:

intransitive verbs

She worries (terribly). = She is a (terrible) worrier.

transitive verbs

(object noun retained in a compounding formation, with tonic placement on the first element of the compound)

He makes trouble (constantly). = He's a ^T(constant) trouble-maker.

- 1.3 an occasional agentive noun, relating to the continuous tenses of an intransitive base verb, with a case-related compound corresponding to a preposition-and-noun phrase:

He is/will be helping tonight. = He is a helper tonight.

T

They are helping in/at the concert. = They are concert-helpers.

- 1.4 a usually generic instrumental noun, relating to the simple tenses of transitive base verbs, with the object noun deleted as understood, unless special contrast is required:

The device strains liquids. = The device is a strainer.

T

The device strains soups. = The device is a soup-strainer.

- 1.5 a compound occupational agentive noun, relating to the simple tenses of intransitive base verbs, case-related and corresponding to a preposition-and-noun phrase:

He works in a factory. = He is a factory-worker.

He lectures in a university. = He is a university lecturer.

She works with a needle. = She is a needle-worker.

- 1.6 a generic or instrumental agentive noun, relating to the simple tenses of intransitive and transitive (compounding) base verbs:

He/It is a person/animal/disease/machine that kills.

= He/It is a killer.

He/It is a person etc that consumes (a lot of/too much) time.

T

= He/It is a time-consumer.

2 +ing as a gerund (verbal noun), two subtypes:

- 2.1 either generic, relating to the simple tenses, or occasional, relating to the continuous tenses, of a base verb:

intransitive verbs

Something { (boils (generally)).
(is boiling (now)). = The action is (called) boiling.

transitive verbs (1)

(object noun retained in a compounding formation, with tonic placement on the first element of the compound)

He { (makes glass (generally)).
(is making glass (now)). = His action is glass-making.

T

transitive verbs (2)

(object noun deleted as understood, unless special contrast is required)

They distil (liquids/whisky). = Their business is distilling.

They distil whisky, not brandy. = Their business is whisky-
distilling, not brandy-
distilling.

- 2.2 as first element in a compound with a purposive pattern, usually generic, relating to the simple tenses of a base verb, with tonic placement on the first element of the compound:

It is a pool for swimming in. = It is a swimming pool.

It is powder for baking with. = It is baking powder.

It is an apple for eating. = It is an eating apple.

It is a factory for making glass.

T

= It is a factory for glass-making.

T
It is a glass-making factory.

- 3 +ing as a present participle, three subtypes:

- 3.1 occasional, relating to the continuous tenses of a base verb:

It is an object which is moving (now). = It is a moving object.

- 3.2 occasional, relating to the continuous tenses, or generic, relating to the simple tenses of intransitive base verbs, compounded with adverbs:

The vehicle { is moving fast (now).
 moves fast (generally). = It is a fast-moving vehicle.

- 3.3 occupational-generic, relating to the simple tenses of a base verb:

intransitive verbs

(retaining the direct object with tonic placement on the first element of the compound)

The girl dances as an occupation. = She is a dancing girl.

The machine adds (by definition). = It is an adding machine.

transitive verbs

(deleting the direct object as understood, unless special contrast is required)

The committee plans policies. = 1 ^T It is a planning committee.
 2 ^T It is a policy-planning committee.

4 +ed as a past participle, four subtypes:

- 4.1 simple perfective, relating to the perfect tenses and the simple tenses as expressing result:

The food has been cooked, and is now cooked. = It is cooked food.

- 4.2 perfective, compounded with an adverb, with tonic placement on the accented syllable of the adverb:

The men have been trained well, and are now well-trained. =
^T
 They are well-trained men.

The men have been trained badly, and are now badly trained. =
^T
 They are badly-trained men.

- 4.3 perfective, compounded with a case-related noun corresponding to a preposition-and-noun phrase, with tonic placement on the accented syllable of the noun:

The mountain is covered with snow. = ^T It is a snow-covered mountain.

The firm is based in London. = ^T It is a London-based firm.

The work is controlled by the government. = ^T It is government-controlled work.

- 4.4 perfective, compounded with the reflexive self-, relating to the perfect tenses and the simple tenses as expressing result:

He has educated himself, and is now self-educated.

= He is a self-educated man.

NOTES1 Orthographic Anomalies

The Latinate holisms using this paradigm are of two kinds: (1) those which take an agentive spelt +er, and are Base Type 2 in Paradigm 7 and (2) those which take an agentive spelt +or, and are Base Type 21 in Paradigm 7. In most accents of standard English the pronunciation is the same for both:

(1) receiver, deceiver, controller, supporter...

(2) selector, inspector, collector, director...

A similar division applies to derived verbs:

(1) classifier, purifier, fertilizer, sterilizer...

(2) dictator, regulator, incubator, percolator...

- 2 Occasionally, such an orthographic difference can be used contrastively:

dictator (= an autocrat)

dictate

dictater (= one who dictates messages)

Contrastive pronunciation of the 'e' and the 'o' in these words can occur to make the distinction clear in speech. Additionally, holistic forms can emphasize the vowel quality in the +er or +or, when an adjective is formed in +ial, with stress on the preceding syllable:

dictator : dictatorial

minister : ministerial

3 Special Patterns for +er

- 3.1 Idiomatization is common with regular formations in +er:

He drinks a lot. = He is a heavy drinker.

She spends a lot. = She is a big spender.

He thinks a lot. = He is a deep thinker.

- 3.2 two minor deverbal patterns are worth noting:

- (1) generic, usually for fruit:

You (can) eat this apple. = It is an eater.

- (2) generic for kinds of transport, especially trains:

You (can) dine in this carriage. = The carriage is a diner.

- 3.3 A listable set (or subset) of +er agentives are either deverbal or denominal:

He farms as an occupation.
He runs a farm. } = He is a farmer.

He golfs as a hobby.
He plays golf. } = He is a golfer.

They ship goods as a business.
They transport goods in ships as a business. } = They are shippers.

A small number of forms such as cricketer and footballer are, however, denominal only.

- 3.4 A generic noun, usually compounded with numbers and describing vehicles, objects or persons, can be formed with +er:

The vehicle weighs three tons. = It is a three-tonner.

The truck has six wheels. = It is a six-wheeler.

The man is six feet tall. = He is a six-footer.

- 3.5 Additionally, for +er there is a place-name pattern, mainly for countries and cities of Germanic provenance and association:

Dublin : Dubliner
Berlin : Berliner
Iceland : Icelander
Maryland : Marylander

- 3.6 Associated with this pattern is a mildly productive pattern for regions:

foreign : foreigner
island : islander
village : villager
lowlands : lowlander
borders : borderer
northern : northerner...
-side : Clydesider, Tynesider, Merseysider...

- 3.7 Similar to +er is +eer, forming nouns denominaly, when the base noun (1) has two or three syllables, and (2) relates to a mechanical object, or a social-commercial activity.

Such words are also usable as verbs with special meanings (to engineer, to electioneer etc) and take a noun in +ing:

engine : engineer : engineering
auction : auctioneer : auctioneering
election : electioneer : electioneering

3.8 There is a mildly productive but small set of words, usually of Greek provenance, which take +ist instead of +er:

type : typist : typing : typing : typed
cycle : cyclist : cycling : cycling : cycled

3.9 Related to this is a set of players of musical instruments:

violin : violinist (but of violin-player)
cello : cellist (but of cello-player)
piano : pianist (but of piano-player)
percussion : percussionist (but of drummer)

4. Special Patterns for +ed

The past participle is paralleled by a denominal adjective, relating to a special correlation of have and be:

simple

The man has a beard. = The man is bearded. He is a bearded man.

The hill has woods on it. = The hill is wooded. It is a wooded hill.

The doctor has (plenty of) experience. = The doctor is experienced.
 He is an experienced doctor.

compounded

The man has a red beard. = The man is red-bearded.
 He is a red-bearded man.

The house has a flat roof. = The house is flat-roofed.
 It is a flat-roofed house.

Paradigm 2 basic deadjectival pattern (Anglo-Saxon provenance)BASE TYPE

Those monosyllabic adjectives, generally of Anglo-Saxon provenance, for which the paraphrases work meaningfully:

quick, hard, white, black...

PARADIGM

<u>a</u>	<u>n</u> 1 <u>+ie</u> (hypocoristic)	
	2 <u>+ness</u> (stative)	
	<u>a</u> <u>+ish</u> (similative) (approximative)	<u>n</u> <u>+ness</u> (stative)
	<u>v</u> <u>+en</u> → PARADIGM 1 (inceptive) (causative)	

EXAMPLESParadigm 2

black
 blackie
 blackness
 blackish
 blackishness
 blacken

Paradigm 1

blackener
 blackening
 blackening
 blackened

DEFINITIVE PARAPHRASES

- 1 +ie, an informal hypocoristic noun, expressing a person's opinion, usually about the nature of someone or something, two subtypes:

1.1 animate

She is, in my opinion, a sweet person. = She's a sweetie.

He is a bad person in that story. = He's a baddie.

1.2 inanimate

It is an old film, compared to others. = It is an oldie.

- 2 +ness, a stative noun:

Redness is the state (or condition) of being red.

Reddishness is the state (or condition) of being reddish.

- 3 +ish, a simulative and approximative adjective, two subtypes:

3.1 simple

The glass is sort of green. = The glass is greenish.

The stone is fairly round. = The stone is roundish.

3.2 compound (for colour adjectives only)

The colour is blue tending towards green/blue containing green/blue with some green in it. = The colour is greenish-blue.
It is a greenish-blue colour.

The cloth is white with some yellow in it. = The cloth is yellowish-white.
It is a yellowish-white cloth.

- 4 +en, an inceptive and/or causative verb (for phonological restrictions, see Note 1):

4.1 inceptive

The sky became/grew dark. = The sky darkened.

The conditions became/grew worse. = The conditions worsened.

4.2 causative

He made the mixture weak(er). = He weakened the mixture.

She made the soup thick(er). = She thickened the soup.

NOTES

- 1 The inceptive-causative verb in +en cannot be formed on bases of the following kinds:
 - 1.1 disyllabic: yellow, hollow...
 - 1.2 ending in a spoken vowel: low, blue...
 - 1.3 ending in a nasal, and often a nasal plus stop consonant: dim, green, blunt...
 - 1.4 ending in an [l] or [r] : little, sour...
 - 1.5 ending in [ʒ] : smooth...
- 2 Two common adjectives ending in a nasal [ŋ] convert to their noun derivative ending in [θ] before adding +en:

long → length: lengthen

strong → strength: strengthen

Additionally, one common adjective ending in a spoken vowel converts in a comparable way to [t] :

high → height: heighten

- 3 A listable and long-established set of adjectives ending in unacceptable sounds take the +en as a prefix en- or em-:

<u>able</u> :	<u>enable</u>	ends in [l]
<u>calm</u> :	<u>encalm</u>	[m]
<u>dear</u> :	<u>endear</u>	[r]
<u>brown</u> :	<u>embrown</u>	[n]
<u>purple</u> :	<u>empurple</u>	[l]

- 4 A listable and long-established set of nouns form a verb in a similar way:

courage: encourage

danger: endanger

balm: embalm

- 5 Three irregular forms have both prefix and suffix:

enliven = to make (something) more lively

enlighten = to make (someone) see the light (=understand)

embolden = to make (someone) bold(er)

- 6 The +en system does not appear to be productive in modern English, and where it is not permitted phonologically, an unchanged adjective converts to a verb:

The clouds became/grew thin. = The clouds thinned.

The flowers became/grew yellow. = The flowers yellowed.

- 7 Adverbial particles are often added to verbs associated with this paradigm, whether en or not:

He made his men as tough as possible. = He toughened his men up.

She made the room really bright with flowers. = She brightened
the room up with flowers.

He made the paint (much) thinner. = He thinned down the paint.

Be a bit calmer, please. = Calm down, please.

Paradigm 3 basic animate denominal pattern (Anglo-Saxon provenance)BASE TYPE

Those simple nouns, of Anglo-Saxon and French provenance, referring to people, for which the paraphrases work meaningfully:

PARADIGM

<u>n</u>	<u>n</u> (stative)	1 <u>+hood</u> (temporal)	
		2 <u>+dom</u> (collective)	
		3 <u>+ship</u> (functional)	
	<u>a</u> (similative)	1 <u>+ly</u> (appreciative)	
		2 <u>+ish</u> (depreciative)	
			<u>n</u> <u>+ness</u> (stative)

EXAMPLES

man
manhood
* mandom
* manship
manly
manliness
mannish
mannishness

king
kinghood
kingdom
kingship
kingly
kingliness
* kingish
* kingishness

queen
queenhood
queendom
* queenship
queenly
queenliness
* queenish
* queenishness

DEFINITIVE PARAPHRASES

- 1 +hood, a stative noun, usually with temporal force:

fatherhood = the state of being a father

manhood = the state of being a man; the time when a man

- 2 +dom, a stative noun, usually suggesting either (1) collectivity, or (2) an area of control:

stardom = the collective state of all stars in the world of entertainment

kingdom = the area of control of a king

- 3 +ship, a stative noun, usually with functional force:

kingship = the state and function of being a king

- 4 +ly, a simlative adjective usually suggesting appreciation:

He is like a man, and I approve of this. = He is manly.

The old man is like a saint, in my opinion. = He is saintly.

- 5 +liness, a stative noun:

Manliness is the state (or condition) of being a man.

- 6 +ish, a simlative adjective usually suggesting depreciation:

The man is/behaves like a fool. = The man is foolish. He is a foolish man.

He behaves like a slave. = He is slavish. He is a slavish fellow.

The man behaves like a woman, and I don't approve of this. = The man is womanish. He is a womanish fellow.

- 7 +ishness, a stative noun:

Foolishness is the state (or condition) of being foolish.

NOTES

- 1 The suffixes +hood, +dom and +ship are not highly productive in modern English, and tend to be mutually exclusive.
- 2 For +hood there are two common countable nouns which are collective and not temporal:

a brotherhood/sisterhood = all the 'brothers/sisters' of a particular kind (usually in a religion or social movement)

- 3 For +ship there is a listable set based on derived nouns in +cr.
Examples:

leadership = the state of being a leader; the ability of a leader

readership = the collectivity of all the readers of a particular magazine etc

- 4 For +ship, there is a listable set of occupational names.
Examples:

a lectureship = the occupation/position of being a lecturer

a readership = the occupation/position of being a reader
(in a British university)

- 5 Related to the preceding is a small set in +ship with the implication of a subsidy paid for a purpose. Examples:

a scholarship = a sum of money granted for purposes of study

a studentship = a sum of money to support a person while a student

- 6 For +ly there is a listable set of adjectives formed on nouns which are not persons. Common members of this set are:

heaven: heavenly

earth: earthly/unearthly

world: worldly

time: timely

death: deathly

dead: deadly

love: lovely

cost: costly

alone: lonely

- 7 For +ish, a possessive element can be found in paraphrases (of Paradigm 4, Note 4):

He had an owlish expression. = He had an expression like an owl's (expression).

- 8 The compound element +like can sometimes be included as a neutral term between +ly and +ish, but does not form a stative noun in +ness:

man: manly: manliness (appreciative)

man: manlike (neutral)

man: mannish: mannishness (depreciative)

- 9 For +ishness, common words convert to an uncountable instance of the state:

Foolishnesses are instances of being foolish.

Paradigm 4. basic inanimate denominal pattern (Anglo-Saxon provenance)BASE TYPES

- 1 Those simple nouns of mainly Anglo-Saxon provenance, mainly monosyllabic, referring to objects and physical situations, for which the paraphrases work meaningfully:

sand, stone, mud, dirt, rust, cheese...

- 2 Comparable disyllabic holisms of any provenance, especially in +er:

paper, powder, sugar, water...

sugar, cabbage, carrot, vinegar...

PARADIGM

<u>n</u>	<u>a</u> + <u>y</u> (attributive) (similative)	<u>n</u> + <u>ness</u> (stative)
----------	--	-------------------------------------

EXAMPLES

sand	powder
sandy	powdery
sandness	powderiness

DEFINITIVE PARAPHRASES

- 1 +y, three subtypes:

- 1.1 an attributive adjective implying have or with:

The water has mud in it. = The water is muddy. It is muddy water.

The beach has sand on it.
 The beach is covered with sand.

} = The beach is sandy. It is a sandy beach.

- 1.2 an attributive adjective implying 'a lot of':

There is a lot of wind today. = It is windy today. It is a windy day.

1.3 a similitive adjective:

The liquid is like oil. = The liquid is oily. It is an oily liquid.

The figure was like a shadow. = The figure was shadowy. It was a shadowy figure.

2 +iness, a stative noun:

Oiliness is the state (or condition) of being oily.

Suniness is the state (or condition) of being sunny.

NOTES

- 1 The suffix +y is generally colloquial and informal, and is particularly well used in domestic language and with children:

It tastes sort of cabbagy. = It tastes a bit like cabbage.

It has a chocolaty taste. = It tastes rather like chocolate.

The dress had a velvety feeling. = The dress felt like velvet.

It's a cottony sort of material. = The material is like cotton.

- 2 Additionally, there is a strong tendency towards emotive self-expression in the use of adjectives formed with +y:

I don't like that man's oily manner.

What a lovely milky complexion she has.

He flashed her a toothy smile.

- 3 In descriptions of colour, +y has a colloquial use similar to +ish in Paradigm 2:

greeny = sort of green

greeny-blue = blue with some green in it/blue tending towards green

- 4 With some animal nouns, usually monosyllabic, +y is used for comparison, and suggests possession (cf Paradigm 3, Note 7):

He has a beard like a goat's (beard). = He has a goaty beard.

She has a manner like a cat's (manner). = She has a catty manner.

- 5 Animal nouns generally use +ly, +ish and +y unsystematically:

lion : lionly

tiger : tigerish

cat : catty

Paradigm 5 denominal in +al (Neo-Latin provenance)BASE TYPES

- 1 Those generally monosyllabic nouns of Latin provenance for which the paraphrases work meaningfully:

form, norm, herb, centre...

- 2 Comparable bound nominal bases of Latin provenance:

reg-, leg-, vit-, mort-...

- 3 Comparable holisms of Latin provenance:

nature, nation, accident, incident...

- 4 Comparable forms requiring specific grapho-phonological adaptations, as in:

sex, text, habit... (adding -u-)

larynx, pharynx (-x → -geal)

vestige (adding -i-)

- 5 Comparable free words and nominal bases containing a lateral [l], usually converting +al to +ar (See Notes 2 to 5):

5.1 free words: lobe, pole, scale, consul...

5.2 bound bases: lun-, sol-, stell-, regul-...

- 6 Nouns with vestigial inflectional endings for the Latin nominative singular (masculine, feminine or neuter), which is dropped:

radius (radi-), continuum (continu-)

- 7 Comparable nouns ending in +y, which in orthography becomes connecting -i-:

industry, colony, family, controversy...

DERIVED TYPES (less likely to exploit the paradigm fully)

- 1 Nouns ending in -nce and -nt, becoming -ntial:

substance, circumstance, existence...

president, resident...

- 2 Nouns and nominal bases ending in:

2.1 -ic(s): logic, music, critic...

radic-

mathematics, ethics, statistics...

- 2.2 -oid(s) adenoid(s), anthropoid...
- 2.3 -or(i-) dictator, pictor-
- 2.4 -ion selection, inflection, direction...
- 2.5 -ment (bound bases mainly)
 monument, experiment, detriment..., government...

PARADIGM

<u>n</u>	<u>a</u> <u>+al</u> (relational) (similative) (attributive)	<u>n</u>	1 <u>+ism</u> (systemic)	PARADIGM 14
		(stative)		
			2 (qualitative)	1 <u>+ity</u> (formal)
				2 <u>+ness</u> (informal)
		<u>v</u> <u>+ize</u> (causative)		PARADIGM 13

EXAMPLESParadigm 5

centre
central
centralism

centrality
centralness
centralize

Paradigm 14

centralist
centralistic

Paradigm 13

centralization

Paradigm 1

centralizer
centralizing
centralizing
centralized

DEFINITIVE PARAPHRASES1 +al, a relational, attributive and/or similitive adjective:

The medicine is obtained from herbs. = The medicine is herbal.
It is herbal medicine.

The colour belongs to nature. = The colour is natural. It is
a natural colour.

2 +alism, see Paradigm 14:

Structuralism is any academic movement which seeks a structural
description or explanation of any phenomena.

Paternalism is the academic name for a social attitude in which
someone behaves in a paternal way towards others, who are sub-
ordinate in some way.

3 +ality, a stative noun, often formal or academic/technical:

Sexuality is the state (or condition) of being sexual/having
sexual attraction.

Normality is the state (or condition) of being normal.

Popularity is the state (or condition) of being popular.

4 +alness, an informal stative noun:

Normalness is the state (or condition) of being normal.

Naturalness is the state (or condition) of being natural.

5 +alize, see Paradigm 13:

He made the situation normal (again). = He normalized the
situation.

They made the system central. = They centralized the system.

He made the place alive again. = He re-vitalized the place.

NOTES1 The suffix +al inherits from Latin two anomalous sets in -n-:

1.1 domestic:

paternal, maternal, fraternal

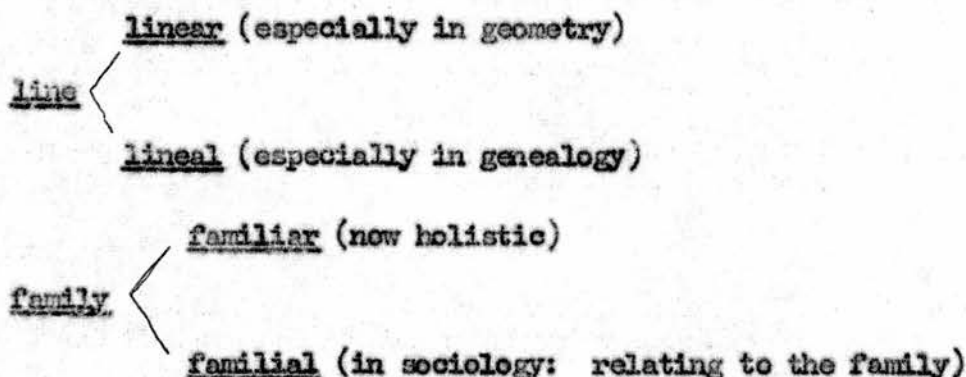
1.2 locational:

exterior: external interior: internal

superior: supernal * eterior: eternal

ulterior: * ulternal inferior/inferno: infernal

- 2 The lateral rule allows a contrast for academic and technical purposes between +ar and +al, as yet only minimally exploited:



- 3 The lateral rule relates particularly in current English to endings in -cle and -gle:

cirole: circular
muscle: muscular
angle: angular

- 4 The lateral rule also applies to endings in -ule, where (1) it is a productive diminutive suffix, and also (2) where it is vestigial:

productive

grain: granule: granular
globe: globule: globular
molecule: molecular

vestigial

tube: tubular
valve: valvular

- 5 There appear to be two sets of exceptions to the lateral rule:

- 5.1 bases containing a lateral but with a final [r] . Such bases take +al:

later-: lateral
flor-: floral
liter-: literal

- 5.2 a listable set of unpredictable exceptions:

colony: colonial
globe: global
palate: palatal
loc-: local
fili-: filial

- 6 Deverbal nouns in +al (such as renew: renewal) have no connection with this paradigm. See Paradigm 7, Note 3.
- 7 Nominalizations (human and non-human) are common with derivatives in adjectival +al, by deletion of an understood noun:

human

He is one of our regular customers/visitors. = He is one of our regulars.

He is a soldier with a regular engagement. He is a regular soldier. = He is a regular.

They are people from the provinces. They are provincial people. = They are provincials.

non-human

She is taking her final examinations. = She is taking her finals.

He had a medical check-up last week. = He had a medical last week.

- 8 Derivations in +ality are generally less formal than parallel derivations in +icity (See Paradigm 9), and +ivity (See Paradigm 7), and many words of this type are in more general use than these others. Compare normality and generality with specificity and conductivity.
- 9 Forms in +ality can also have a countable use, as an instance of their uncountable state:

Formalities are instances of formality.

Irregularities are instances of irregularity.

Paradigm 6 denominal in +an (Neo-Latin provenance)BASE TYPES

- 1 Those place-names as listed below, for which the paraphrases work meaningfully:

- 1 Ending in simple -a:

Cuba, Alaska, Africa, Asia...

- 2 Ending in -ia:

India, Rumania, Indonesia, Arabia...

- 3 Ending in -y, with an orthographic change to -i- and an accent shift one syllable towards the suffix:

Italy, Sicily, Burgundy, Hungary...

- 4 Ending in a consonant and adding +ian:

Brazil, Iran, Egypt, Ecuador...

- 5 A listable set of special forms with various spoken-vowel endings, converting in various irregular ways to +an. Examples:

Canada: Canadian Peru: Peruvian Mexico: Mexican

Norway: Norwegian Glasgow: Glaswegian Manchester: Mancunian

PARADIGM

<u>n</u>	<u>a</u> + <u>an</u> (relational)	<u>n</u> (stative)	1 + <u>ism</u> Paradigm 14 (systemic)
			2 + <u>ness</u> (qualitative)
		<u>v</u> + <u>ize</u> Paradigm 13 (causative)	

EXAMPLESParadigm 6

India
Indian
Indianism

Paradigm 14

Indianist
Indianistic

Paradigm 13

Indianization

Paradigm 1

Indianizer
Indianizing
Indianizing
Indianized

Indianize

DEFINITIVE PARAPHRASES

- 1 +an, a relational adjective for place-names:

The man is from America. = The man is American.

The plant belongs to India. = The plant is Indian. It is an Indian plant.

- 2 +anism, see Paradigm 14 for use as a stative systemic noun, and Paradigm 15 for use as a type noun:

stative

Indianism is the cultural attitude which elevates everything Indian/is the political belief which places India first.

type

An Indianism is a linguistic or cultural peculiarity of India/the Indians.

- 3 +anness, a stative noun:

Canadianness is the state (or condition) of being Canadian.

- 4 +anize, see Paradigm 13:

He made the Service Indian. = He Indianized the Service.

NOTES

- 1 There are two noun bases which might be assumed to take +al but take +an:

hum-: human urb-: urban

- 2 This set has a non-productive parallel in +ane:

human: humane (using the paradigm)

urban: urbane (not using the paradigm)

- 3 Only this set has a stative noun in +anity, accent on -an-:

human/humane: humanity urbane: urbanity

- 4 Human also has a unique derivational system:

human: humanity: humanitarian: humanitarianism

- 5 A productive academic use of +an exists, but does not normally exploit the paradigm:

Sioux: Siouan Aztec: Aztecan

- 6 The form +ian may be added to personal surnames, usually those with two syllables, and may exploit the paradigm. The addition of the suffix attracts the accent one syllable towards the suffix and may change the vowel quality in the accented syllable:

Bacon: Baconian Johnson: Johnsonian

Newton: Newtonian Adler: Adlerian

Freud: Freudian Jung: Jungian

- 7 There are three (or four) place-names in +y which form their adjective by deletion and/or adaptation:

Germany: German

Tuscany: Tuscan

Normandy: Norman

(Brittany: Breton)

Paradigm 7 deverbal in +ion/ive (Neo-Latin provenance)BASE TYPES

- 1 These preponderantly disyllabic holistic verbs of Latin provenance for which the paraphrases work meaningfully (See also Paradigm 1):

select, inspect, reflect, predict, act...

- 2 Such verbs for which special grapho-phonological adaptations are necessary:

receive (recept-), evade (evas-), divide (divis-), erode (eros-)
declare (declarat-), compare (comparat-), add (addit-)...

DERIVED TYPE (less likely to exploit the paradigm fully)

Di- and trisyllabic verbs in +ate:

negate, indicate, nominate, investigate...

PARADIGM

v	n <u>+ion</u> (actional)	a <u>+al</u> (relational) (similative) (attributive)
	a <u>+ive</u> (purposive)	n (stative) 1 <u>+ism</u> → PARADIGM 14 2 <u>+ity</u> (formal) 3 <u>+ness</u> (informal)
→ PARADIGM 1		

EXAMPLESParadigm 7

collect
collection
collectional
collective
collectivism

collectivity
collectiveness

Paradigm 14

collectivist
collectivistic

Paradigm 1

collector
collecting
collecting
collected

DEFINITIVE PARAPHRASES

- 1 +ion, an actional noun, which can be expressed actively or passively:

Collection is the action, business, process etc of collecting or being collected.

Inspection is the action, business, process etc of inspecting or being inspected.

Investigation is the action, business, process etc of investigating or being investigated.

- 2 +ional, see Paradigm 5, a relational, simulative and/or attributive adjective:

The procedure relates to selection. = The procedure is selectional.
It is selectional procedure.

- 3 +ive, a purposive adjective:

The purpose of the procedure is selection/to select. =
The procedure is selective. It is selective procedure.

- 4 +ivism, see Paradigm 14, a systemic stative noun:

Negativism is the state (or condition) of being negative, defined as a social attitude.

- 5 +ivity, a formal stative noun:

Selectivity is the state (or condition) of being selective.

- 6 +iveness, an informal stative noun:

Selectiveness is the state (or condition) of being selective.

- 7 for other forms of the verb see Paradigm 1.

NOTES

- 1 In the terminology of grammar, +ive can be added, also with a purposive force, to nouns:

stative = for the purpose of expressing a state (or condition)

purposive = for the purpose of expressing a purpose

- 2 The major grapho-phonological adaptations of bases for the addition of +ion are shown in the sets below, two examples per set:

2.1	<u>express:</u>	<u>expression</u>	<u>suppress:</u>	<u>suppression</u>
2.2	<u>concede:</u>	<u>concession</u>	<u>recede:</u>	<u>recession</u>
2.3	<u>suggest:</u>	<u>suggestion</u>	<u>digest:</u>	<u>digestion</u>
2.4	<u>admit:</u>	<u>admission</u>	<u>permit:</u>	<u>permission</u>
2.5	<u>delete:</u>	<u>deletion</u>	<u>complete:</u>	<u>completion</u>
2.6	<u>assert:</u>	<u>assertion</u>	<u>exert:</u>	<u>exertion</u>
2.7	<u>extend:</u>	<u>extension</u>	<u>expand:</u>	<u>expansion</u>
2.8	<u>intend:</u>	<u>intention</u>	<u>attend:</u>	<u>attention</u>
2.9	<u>convert:</u>	<u>conversion</u>	<u>subvert:</u>	<u>subversion</u>
2.10	<u>evolve:</u>	<u>evolution</u>	<u>dissolve:</u>	<u>dissolution</u>
2.11	<u>attribute:</u>	<u>attribution</u>	<u>contribute:</u>	<u>contribution</u>
2.12	<u>propel:</u>	<u>propulsion</u>	<u>compel:</u>	<u>compulsion</u>
2.13	<u>inscribe:</u>	<u>inscription</u>	<u>prescribe:</u>	<u>prescription</u>
2.14	<u>erupt:</u>	<u>eruption</u>	<u>corrupt:</u>	<u>corruption</u>
2.15	<u>resume:</u>	<u>resumption</u>	<u>consume:</u>	<u>consumption</u>
2.16	<u>extract:</u>	<u>extraction</u>	<u>connect:</u>	<u>connection</u>
2.17	<u>erode:</u>	<u>erosion</u>	<u>evade:</u>	<u>evasion</u>
2.18	<u>decide:</u>	<u>decision</u>	<u>provide:</u>	<u>provision</u>
2.19	<u>include:</u>	<u>inclusion</u>	<u>delude:</u>	<u>delusion</u>
2.20	<u>prohibit:</u>	<u>prohibition</u>	<u>inhibit:</u>	<u>inhibition</u>
2.21	<u>allege:</u>	<u>allegation</u>	<u>oblige:</u>	<u>obligation</u>
2.22	<u>reserve:</u>	<u>reservation</u>	<u>conserve:</u>	<u>conservation</u>
2.23	<u>indicate:</u>	<u>indication</u>	<u>regulate:</u>	<u>regulation</u>
2.24	<u>classify:</u>	<u>classification</u>	<u>purify:</u>	<u>purification</u> (See Paradigm 12)
2.25	<u>centralize:</u>	<u>centralization</u>	<u>pasteurize:</u>	<u>pasteurization</u> (See Paradigm 13)

- 3 There are three alternative suffixes to +ion for nominalization from Latinate verbs. They are:

- 3.1 +ment, formed typically on bases of from one to three syllables, but ending in sounds unsuitable for adaptation as above to +ion, and often with special phonological patterns:

unsuitablepay: paymentjudge: judg(e)mentconceal: concealmentgovern: governmentargue: argumentSpecial patternsad, ap _____advance, appoint, appease...be _____bewilder, beguile, bereave...en _____entice, enlighten, encourage, enlarge...em _____embarrass, embitter...in _____indict...im _____imprison, improve, impoverish...

- 3.2 +al, formed typically on bases which have two syllables but cannot relate to +ion, and/or relate to acts of will, planning, the law etc:

betray: betrayalacquit: acquittalrefuse: refusalremove: removalrenew: renewalarrive: arrival

- 3.3 +ure, a very few special words and bases:

depart: departurefail: failureenclose: enclosureforfeit: forfeiturefract-: fracturepict-: picture

- 3.4 +y, a very few special words:

recover: recoverydiscover: discoverydeliver: delivery

- 4 There are possibilities of meaningful contrast between forms in +ion and the others, occasionally exploited:

induceinduction (in thinking)inducement (eg bribing, incentives)

- 5 The actional noun in +ion etc often has a resultative or instantial use:

Collections are instances of or the results of collection.

Recoveries are instances of or the results of recovery.

Paradigm 8 restricted denominals (Neo-Latin provenance)BASE TYPES

- 1 Listable sets, as exemplified below, of bound nominal bases of Latin provenance whose meaning can often be expressed by glosses (= relexification into free words of English):

- 1.1 +id rigid, lucid, placid, torpid...
 1.2 +ile fertile, mobile, sterile, fragile...
 1.3 +ine canine, feline, aquiline, masculine...

- 2 A listable set of nouns and bound bases in +ous:

- 2.1 nouns: pomp, bulb, nerve, fibre, nebul(a)...
 2.2 other nominal bases: visc-, curi-, dubi-, vacu-...

PARADIGM

<u>n</u>	<u>a</u> <u>+id</u> <u>+ile</u> <u>+ine</u> <u>+ous</u> (relational) (attributive) (similative)	<u>n</u> (stative)	1 <u>idity</u> <u>ility</u> <u>inity</u> <u>osity</u> (formal)
			2 <u>+ness</u> (informal)

EXAMPLES

luc- lucid lucidity lucidness	sal- saline salinity salineness	mob- mobile mobility mobileness
pomp pompous pomposity pompousness	visc- viscous viscosity viscousness	

DEFINITIVE GLOSSES AND PARAPHRASES

- 1 +id, +ile, +ine, +ous, a relational, attributive and/or similative adjective, often explicable by gloss rather than paraphrase, and therefore tending to be holistic:

He thinks clearly. = He is a lucid thinker.

The material does not bend. = It is rigid material.

You can grow things easily in this soil. = It is fertile soil.

This kind of home is built to move. = It is a mobile home.

The animal has a manner rather like a cat's. = It has a feline manner.

These teeth are like a dog's. = These are canine teeth.

That man is full of his own importance. = He is a pompous fellow.

The effect is like a cloud. = The effect is nebulous.

2 +ity, a stative noun, formal and often academic/technical:

Lucidity is the state (or condition) of being lucid, formally stated.

Viscosity is the state (or condition) of being viscous, technically described.

3 +ness, an informal stative noun:

Lucidness is the state (or condition) of being lucid.

Viscousness is the state (or condition) of being viscous.

NOTES

- 1 The bulk of the adjectives exploiting this paradigm are holisms and therefore analysis of their elements is an analysis of root and suffix rather than base and suffix. Analysis ranges along a continuum from the useless to the useful:

useless

placid = * inclined to please

viscous = * relating to mistletoe, birdlime

slightly more useful

fertile = * able to bear, yield

lucid = * giving off light

useful

mobile = able to move

saline = relating to or containing salt

highly useful

feminine = like or relating to a woman (femin-)

pompous = full of pomp

- 2 The forms +ity and +ness appear to be in competition, the formal +ity being less popular:

nervous: nervousness but not *nervosity

bulbous: bulbousness but not *bulbosity

However, certain long-established forms favour the +ity, with the suggestion of a difference in meaning between +ity and +ness:

<u>curious</u>	<	<u>curiosity</u>	1 the state of being curious
			2 the inclination to be curious
			<u>curiousness</u>

- 3 There is a listable set of bases which take +ous via a connecting element -aci- and exploit the paradigm as follows:

ten-: tenaci-: tenacious: tenacity: tenaciousness

vor-: voraci-: voracious: voracity: voraciousness

Paradigm 9 denominal in +ic (Neo-Latin provenance)BASE TYPES

- 1 Those simple nouns of Greek provenance, mainly disyllabic, for which the paraphrases work meaningfully:

angel, nomad, icon, cycle, tone...

- 2 Comparable holistic nouns of Greek provenance:

atom, symbol...

- 3 Comparable simple or holistic nouns of Greek provenance ending in -ia and -y, where these are dropped:

3.1 simple: mania, phobia, irony, history, harmony...

3.2 holistic: allergy, category, symmetry...

- 4 Comparable bound bases of Latin or Greek provenance:

civ-, publ-, graph-, ethn-, didact-, gastr-...

- 5 Listable nouns or nominal bases of Greek provenance ending in -m(a) and taking -(a)t- as a connecting element:

5.1 -ma: dogma, charisma, aroma...

5.2 -m: axiom, idiom, system, phlegm, emblem...

5.3 comparable bases: chrom-, pragm-...

- 6 Listable nouns and nominal bases forming in -e(u)tic:

6.1 -etic: phon-, athl-, aesth-...

6.2 -eutic: therapy, pharmacy...

DERIVED TYPES (less likely to exploit the paradigm fully)

- 1 Derived on a Neo-Latin, especially Greek, basis:

dyslexia, hypertrophy, aphasia...

- 2 Derived in relation to a form compounded on Neo-Latin, especially Greek, principles (See Notes 5 and 10):

democracy: democratic

photography: photographic

arthritis: arthritic

neurosis: neurotic

- 2 +al, basically identical with +ic, but usually suggesting a generalized quality, often in contrast with +ic (See Note 11):

The movement relates in a general way to cycles/ = The movement is
is in a general sense like a cycle/has, in cyclical. It
general terms, the attributes of a cycle. is a cyclical
movement.

- 3 +icity, attracting the accent to the -ic- element, a formal or academic/technical stative noun:

Tonicity is the state (or condition) of being tonic (as defined in phonetics).

Historicity is the state (or condition) of being historic (as understood by historians).

- 4 +iness, an informal stative noun:

Toxicness is the state (or condition) of being toxic, understood in a non-technical sense.

NOTES

- 1 The accent shift associated with +ic is very regular:

'atom a'tomic
'athlete a'thletic

but there are four common exceptions:

'Catholic, 'Arabic, a'rithmetic, 'politics

- 2 Nouns of Latin and Greek provenance with vestigial inflectional endings for the nominative singular, especially masculine -us and -es, and including proper names, drop the ending and take +ic, but do not usually exploit the paradigm:

citrus (citr-) nucleus (nucle-)
Olympus (Olymp-) Socrates (Socrat-)

- 3 Type, group and place names, regardless of provenance, and usually for purposes of academic or technical terminology take +ic, but also do not exploit the paradigm (See also Note 14):

3.1 by simple addition: Iceland, Lapland, Sanskrit, Czech...

3.2 by dropping a final -a especially if of Sanskrit provenance:

Buddha, Veda, karma, yoga...

- 4 Personal surnames and group names ending in a nasal and preponderantly disyllabic in form, and usually for academic purposes can take +ic, and do not exploit the paradigm:

Byron, Milton, Teuton, Titan, German...

- 5 Compounds of Greek provenance or type generally form their adjectives as described here, with these exceptions:

5.1 -graphy forms have a clearcut +ic/ical distinction, as above:

photography: photographic: photographical

5.2 -logy forms take only -ical:

zoology: *zoologic: zoological

5.3 -cracy forms take only +ic:

democracy: democratic: *democratical

- 6 A listable set of long-established words ending in -a reject the rule that the -a must be dropped, and take the accent on the -a, with a change of vowel quality [ə] to [e] :

algebra: algebraic

stanza: stanzaic

formula: formulaic

NOTE prosaic is now a holism, only historically linked with prose.

- 7 To these can be added some classical personal and place names.
Examples:

Ptolemy: Ptolenaic

Moses: Mosaic

Hebrew: Hebraic

Rome: Romaic (= Byzantine)

Aram: Aramaic

Chaldea: Chaldaic (contrasting with Chaldean)

Judea: Judaic (contrasting with Judean)


- 8 Three nouns used in poetic scansion are irregular:

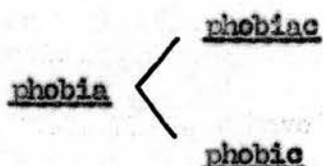
strophe: strophaic

trochee: trochaic

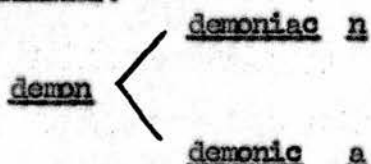
spondee: spondaic

- 9 With medico-pathological nouns, there is competition (and potential contrast) between +ic and +iac:

mania  maniac, maniacal
manic



There is a tendency to use +iac forms as personal nouns. Compare the similar:



In some long-established terms, however, +iac is the norm, whether adjectival or nominal:

cardiac, ammoniac

- 10 There are listable sets of nouns, usually compound, of Greek provenance or type, with special grapho-phonological changes in derivation. These do not, however, normally exploit the paradigm:

10.1 -se ellipse: elliptic: elliptical
 synopse: synoptic

10.2 -sy epilepsy: epileptic
 catalepsy: cataleptic

10.3 -sia dyspensia: dyspeptic
 eupensia: eupeptic

10.4 -sis syllipsis: syllaptic
 paralysis: paralytic
 catharsis: cathartic

10.5 -osis otic osmosis: osmotic
 neurosis: neurotic

10.6 -itis itic arthritis: arthritic
 hepatitis: hepatitic

10.7 -ite Semite: Semitic
 Hamite: Hamitic

- 11 Although it is difficult in some instances to find glosses which indicate that contrastive -ic and -ical are in a relationship of special to general, the following list helps to indicate such a tendency:

BASE	SPECIAL	GENERAL
poet	poetic	poetical
history	historic	historical
allegory	allegoric	allegorical
metaphor	metaphoric	metaphorical
irony	ironic	ironical
cycle	cyclic	cyclical
electr-on	electric	electrical
economy	economic	economical
astronomy	astronomic	astronomical
philosophy	philosophic	philosophical
geography	geographic	geographical
botany	botanic	botanical
academy	academic	academical

- 12 Occasional use is made of +ic, added for academic purposes to adjectival bases in the suffixes +al (+ar) and +an, to make 'double adjectives', usually relational:

velum: velar: velaric (= of the velar type)

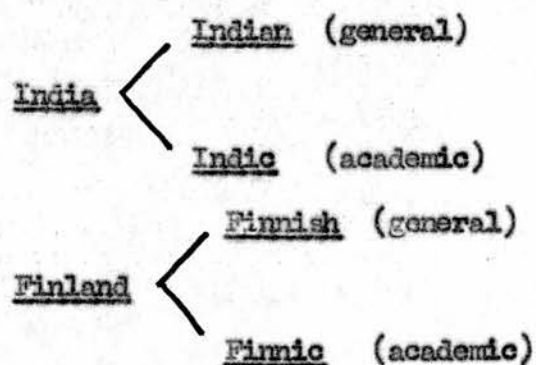
voice: vocal: vocalic (= of the vowel type)

Rome: Roman: Romeric (= of the Roman type)

- 13 +ism is occasionally added to +ic for a systemic stative noun:

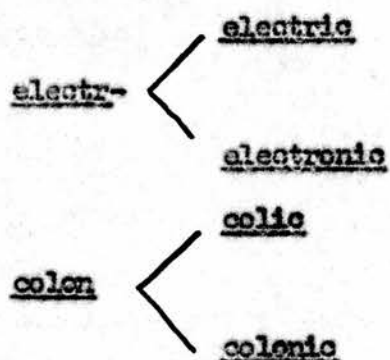
Esotericism is the attitude of life which favours being esoteric.

- 14 The academic use of +ic with type, group and place names contrasts for academic purposes with +an and +ish:



- 15 A special +onic variant seems to be gaining popularity for two reasons:

15.1 contrastive with existing adjectives in +ic:



15.2 to add grapho-phonological bulk to a form that would otherwise be unusable:

bi-: bionic (not *bic)
psi-: psionic (not *psic)

Paradigm 10 nominalizing in +ic(s) (Neo-Latin provenance)BASE TYPE

Those bound nominal bases of Greek and Latin provenance for which the paraphrases work meaningfully, and for which +ic (see Paradigm 9) is the typical adjectival suffix:

eth-, graph-, ceram-, phys-...

DERIVED TYPE (less likely to exploit the paradigm fully)

Bound derivatives in -ist- of both Greek and Latin provenance:

heurst-, statist-, linguist-...

PARADIGM

<u>n</u>	<u>n</u> + <u>ics</u> (type: abstract)	<u>n</u> (type: human)	
		1 + <u>ist</u>	
		2 + <u>ion</u>	
		<u>a</u> → <u>ic</u> (relational) (attributive) (SPECIALIZING)	<u>a</u> + <u>al</u> (relational) (attributive) (GENERALIZING)
		<u>v</u> + <u>ize</u> → PARADIGM 13 (causative) (performative)	

EXAMPLESParadigm 10

ethics
ethicist
ethician
ethic
ethical
ethicize

Paradigm 13

* ethicization

Paradigm 1

* ethicizer
ethicizing
ethicizing
ethicized

Paradigm 10

politics
* politician
politician
politio
political
politicize

Paradigm 11

* politicization

Paradigm 1

* politicizer
politicizing
politicizing
politicized

DEFINING PARATHESES

- 1 +ics a noun expressing an abstract type or system:

Graphics is (1) that aspect of engineering which concerns the actual drawing (= graph-) of plans (2) that branch of mathematics which is concerned with calculating by means of diagrams.

Ceramics is that branch of art which is concerned with fired pottery (= ceram-).

Statistics is that branch of science/mathematics which deals with the manipulation of numerical data (= statist-).

- 2 +icist/+ician (See also Note 2) a personal noun:

He studies physics. = He is a physicist.

He studies genetics. = He is a geneticist.

He studies statistics. = He is a statistician.

He studies mathematics. = He is a mathematician.

- 3 → ic and → ical, a relational or attributive adjective (See Paradigm 9, Note 11):

The information relates to semantics./The information has the attributes of semantics. = The information is semantic. It is semantic information.

The work relates to politics. = The work is political. It is political work.

- 4 +icize, see Paradigm 13.

He made the problem political. = He politicized the problem.

NOTES

1. This paradigm is restricted to academic usage, and is seldom if ever fully exploited.

- 2 In the personal noun additions +ist and +ian there is usually competition, not contrast, depending apparently on ease of pronunciation:

genetics: geneticist not *genetician

statistics: statistician not *statisticist

Regular formation is not predictable for words that might exploit the paradigm in future. Additionally, preference for simpler terms (even if these are ambiguous) militates against the full use of this paradigm:

linguistics: linguist not usually linguistician
or *linguisticist

The incidence of physicist and physician suggests, however, that contrastive possibilities are latent in the pair.

- 3 Similarly, for this paradigm the contrasts or preferences between +ic and +ical are not as sharp as in Paradigm 9. The suffix +ic, is however, a back-formation (+ics → ic) and when a countable noun occurs in +ic the adjective in +ical is preferred:

ethics: an ethic: ethical

statistics: a statistic: statistical

polemics: a polemic: polemical

- 4 There is a listable set of long-established uncountable nouns in +ic which use the paradigm, and which always forms its adjectives in +ical:

music: musical

magic: magical

arithmetic: arithmetical

- 5 There is also a listable set of countable nouns in +ic (human and non-human) which does not use the paradigm but is cognate:

5.1 human: a critic: critical: criticize: criticism

5.2 non-human: a topic: topical: topicalize

- 6 For forms in +ian there is a possible extension, seldom exploited, to a stative noun in +ship:

musician: musicianship (= the state or condition of being a musician)

Paradigm 11 denominal in +oid (Neo-Latin provenance)BASE TYPE

- 1 Those bound nominal bases of Greek provenance for which the paraphrases work meaningfully:

anthrop-, pithec-, aster-, lith-...

- 2 Comparable but infrequent nouns of Greek provenance:

cube, globe...

- 3 Comparable nouns of Greek provenance with a vestigial Latin inflectional nominative ending, which must be deleted:

rhombus (rhomb-), trapezium (trapez-)...

- 4 Specific academic type and group names, regardless of provenance:

Mongol, Caucasus (Caucas-), Negro (Negr-), Australia (Austral-)...

PARADIGM

<u>n</u>	<u>a +oid</u> (similative)	<u>n +ism</u> (stative) (systemic)
----------	-------------------------------	--

EXAMPLES

anthrop-
anthropoid
anthropoidism

DEFINITIVE PARAPHRASES

- 1 +oid, a similative adjective:

The object is like a cube. = The object is cuboid.

The beast is man-like. = The beast is anthropoid.
(anthrop- = man)

The object in scientific terms is = The object is lithoid.
like a stone/stone-like.

- 2 +oidism, a systemic stative noun:

Adenoidism is a medical condition of the adenoid tissue.

NOTES

- 1 There is a regular conversion for +oid adjectives into countable nouns:

An anthropoid creature is an anthropoid.

A lithoid object is a lithoid.

- 2 This nominalizing conversion has an adjective in +al, which does not, however, normally exploit the +al paradigm:

anthropoid: anthropoidal

adenoid: adenoidal

- 3 Additionally, the nominal conversion in the plural can be used for medical descriptions:

adenoids = enlargement of glandular (aden-) tissue at the back of the nose

haemorrhoids = piles

Paradigm 12 denominal and deadjectival +ify (Neo-Latin provenance)BASE TYPES

- 1 Those mainly monosyllabic simple nouns and adjectives of Latin or Greek provenance for which the paraphrases work meaningfully:

1.1 nouns: class, note, verse, code, person...

1.2 adjectives: simple, pure, false, null, diverse, intense...

- 2 Comparable nouns with a vestigial nominative singular (mainly neuter, Latin or Greek), which must be deleted:

calcium (calc-), stratum (strat-), ammonium (ammoni-),
electron (electr-)...

- 3 Comparable nouns ending in a written -y, which must be deleted (or assumed into +ify as -i-):

beauty (beaut-), glory (glor-), mercury (mercur-)...

- 4 Any comparable noun or adjective where an ortho-phonological adaptation (usually in vowel quality) takes place:

4.1 nouns: type (typ-), metre (metr-), mode (mod-),
sign (sign-), saint (sanct-), fruit (fruct-),
peace (pac-)...

4.2 adjectives: vile (vil-), clear (clar-)...

- 5 Comparable bound bases:

5.1 nouns: petr-, vit-, de-, oss-, ram-...

5.2 adjectives: magn-, quant-, qual-, ver-, rect-, fort-...

PARADIGM

<u>n/a</u>	<u>v</u> + <u>ify</u>	<u>n</u> → <u>ification</u>	<u>a</u> 1 + <u>al</u>
	(inceptive)	(actional)	(relational)
	(causative)	(resultative)	
	(factitive)	(type)	2 → <u>ificatory</u>
			(purposive)
	→ PARADIGM 1		

EXAMPLESParadigm 12

pure
purify

purification
purificational
purificatory

Paradigm 1

purifier
purifying
purifying
purified

DEFINITIVE PARAPHRASES1 +ify, three kinds of verb, often with relexification:

1.1 an intransitive inceptive verb:

The forest turned into stone. = The forest petrified.

1.2 an intransitive factitive verb:

He makes verses. = He versifies.

1.3 a transitive causative verb:

He made the water pure. = He purified the water.

She made her thoughts clear(er). = She clarified her thoughts.

They made the microbe (appear) big. = They magnified the
microbe.

He arranged the material in(to) = He classified the material.
classes.

They transformed the substance = They vitrified the substance.
into glass.

They treated the substance with = They zincified the substance.
zinc.

2 +ification, with an accent on -cat-, two kinds of noun:

2.1 an actional or resultative noun:

Purification is the {action} of {purifying something.
(result) (making something pure.

Clarification is the {action} of {clarifying something.
(result) (making something clear.

Magnification is the $\begin{cases} \text{(action)} \\ \text{(result)} \end{cases}$ of $\begin{cases} \text{(magnifying something.} \\ \text{(making something (seem) big(ger).} \end{cases}$

Zincification is the $\begin{cases} \text{(action)} \\ \text{(result)} \end{cases}$ of $\begin{cases} \text{(zincifying something.} \\ \text{(treating something with zinc.} \end{cases}$

- 2.2 an optional countable type noun for instances or results of an action:

A fortification is an instance of fortification.

Clarifications are instances of clarification.

A classification is the result of classification.

- 3 +cational, with an accent on -cat-, a relational adjective:

The rite relates to purification. = The rite is purificational.
It is a purificational rite.

- 4 +catory, with an accent on -cat-, a purposive adjective:

The rite is for the purpose of purification/of purifying. = The rite is purificatory.
It is a purificatory rite.

NOTES

- 1 There is a listable set of long-established and commonly occurring derivatives in +ify which do not conform to the paraphrase options. They are best accounted for with semi-paraphrases and glosses:

1.1 semi-paraphrases

He proved/showed something to be true. = He verified it.

He made something less dense. = He rarefied it.

He delineated something clearly. = He specified it.

He adapted something. = He modified it.

He tortured/killed someone (on a cross). = He crucified him.

He made the sound louder. = He amplified it.

He expressed the thing in exact amounts. = He quantified it.

The experience humiliated us. = We were mortified.

He was suitable for the job. = He qualified.

1.2 glosses

I informed him. = I notified him.

He limited it. = He qualified it.

- 2 Two minor patterns for +ify appear to be currently non-productive:

2.1 with +ful

He made it	peaceful. beautiful. fruitful.	=	He	pacified beautified fructified	it.
------------	--------------------------------------	---	----	--------------------------------------	-----

2.2

He It	served as	a	type sign example	for	something.
		an embodiment of			

He It	typified signified exemplified personified	something.
----------	---	------------

- 3 A non-productive listable set of forms derived from bound bases of Latin provenance takes a causative written as +efy, indistinguishable in speech from +ify. Such forms tend to have an adjective in +id (See Paradigm 8). Examples:

tum- (swollen) = tumid = tumefy
pingu- (fat) = pinguid = pinguefy
putr- (decay) = putrid = putrefy

- 4 This pattern seems to have been replaced by +idify. Compare liquefy and solidify, humidify.

- 5 Forms in +efy take their stative noun in →efaction:

liquefy: liquefaction
putrefy: putrefaction

- 6 They also have an additional instrumental and type noun in →efacient, with an optional adjectival use:

A liquefacient is anything which causes liquid to form, and has a liquefacient purpose.

- 7 This use has extended separately into medical terminology as a causative instrumental compounding element:

an abortifacient = an agent which causes abortion

- 8 A similar instrumental noun can occur in the +ify system for similar academic/technical purposes:

clear: clarify: a clarificant

- 9 There is a listable subset of adjectives in +ific of two types:

- 9.1 forming on a verb in +ify

horrify: horrific

terrify: terrific

- 9.2 having no comparable verb and probably best analysed as a compounding element +fic (=making, causing):

* scientific: scientific

- 10 See also Paradigm 13, Note 2 for a general comparison of +ify and +ize.

Paradigm 13 denominal and deadjectival +ize (Neo-Latin provenance)BASE TYPE

- 1 Those mainly disyllabic often holistic nouns of Latin and Greek provenance for which all or any of the paraphrases work meaningfully:

atom, itan, hybrid, symbol...

- 2 Comparable nouns with a vestigial inflectional ending of any kind, which must be deleted (See Note 5):

formula (formul-), silicon (silic-), minimum (minim-),
maximum (maxim-), epitome (epitom-)...

- 3 Comparable nouns ending in -y, which must be deleted (or assumed as -i- into +ize):

subsidy (subsid-), sodomy (sodom-), deputy (deput-)...

- 4 Comparable bound bases:

femin-, pulver-, cauter-, ostrac-...

- 5 Personal surnames, mainly disyllabic and regardless of provenance:

Mesmer, Stalin, Milton, Bowdler, Macadam...

- 6 Group and type names, mainly disyllabic and regardless of provenance:

Creole, Hellene, Hindu, Islam, Talmud...

DERIVED TYPES

Adjectives and nouns derivable from words with the following suffixes, whether functional or parts of holisms:

- 1 +al brutal, vandal...
- 2 +an human, African...
- 3 +ar regular, popular...
- 4 +ary military...
- 5 +er tender, Bowdler...
- 6 +ian Prussian, Indian...
- 7 +ic public, critic...
- 8 +ical topical, radical...
- 9 +id liquid...
- 10 +il(e) fossil, fertile...
- 11 +in(e) nicotine...
- 12 +ion union...
- 13 +ional regional, national...
- 14 +ive subjective, objective...
- 15 +mental compartmental...

PARADIGM

<u>n/a</u>	<u>v +ize</u> (inceptive) (performative) (factitive) (causative)	<u>n +ation</u> (actional) (resultative) (type)
	→ PARADIGM 1	

EXAMPLESParadigm 13

atom
atomize

atomization

Paradigm 1

atomizer
atomizing
atomizing
atomized

DEFINITIVE PARAPHRASES1 +ize, four kinds of verb:

1.1 an intransitive inceptive verb:

The spirit became material. = It materialized.

The liquid turned into atoms. = It atomized.
(= droplets)

1.2 an intransitive performative verb:

He served as a deputy. = He deputized.

1.3 an intransitive factitive verb:

He made sermons. = He sermonized.

1.4 a transitive causative verb:

He made their marriage regular. = He regularized their marriage.

He formed the men into a union. = He unionized the men.

He treated the material with sulphur. = He sulphurized the material.

He treated the play as Bowdler did,
by removing the supposedly indecent elements. = He bowdlerized the play.

2 +ization, with an accent on -sat-, two kinds of noun:

2.1 an actional or resultative noun:

Atomization is the $\left\{ \begin{array}{l} \text{(action)} \\ \text{(result)} \end{array} \right\}$ of $\left\{ \begin{array}{l} \text{(atomizing something.} \\ \text{(turning something into atoms.} \end{array} \right.$

Centralization is the $\left\{ \begin{array}{l} \text{(action)} \\ \text{(result)} \end{array} \right\}$ of $\left\{ \begin{array}{l} \text{(centralizing something.} \\ \text{(making something (more) central.} \end{array} \right.$

2.2 an optional countable type noun for instances or results of an action:

Organizations are $\left\{ \begin{array}{l} \text{(instances)} \\ \text{(the result)} \end{array} \right\}$ of organization.

NOTES

1 The spellings +ize and +ise are both common. The first is preferred in the United States and by certain British publishers (for example, Oxford University Press), but +ise is considered British in the United States.

2 The suffix +ify tends to be used with monosyllabic bases, while +ize is used with disyllabic bases. The result in both cases is a trisyllabic derivative with the same rhythm and stress:

'purify 'organize

3 Although tradition is strong in restricting both +ify and +ize to bases of Neo-Latin provenance, they are both used casually with bases of Anglo-Saxon provenance or type which have some kind of phonological similarity to the base or derived types:

<u>womanize</u>	<u>dandify</u>
<u>gardenize</u>	<u>speechify</u>
<u>slenderize</u>	<u>Frenchify</u>
<u>tenderize</u>	<u>Anglify</u>
<u>goblinize</u>	<u>sissify</u>

4 Casual coinages may or may not follow the paraphrases closely or at all:

4.1 following the paraphrases speechify = to make speeches

4.2 not following the paraphrases womanize = to seek sex with many women

5 The inflectional ending -um (Latin neuter nominative) does not follow the deletion rule when preceded by a vowel:

barium: bariumize

vacuum: vacuumize

radium: radiumize

medium: mediumize

- 6 Although most forms taking +ion can form a relational adjective in +ional, this paradigm does not, except for:

organize: organization: organizational

- 7 See Paradigm 14. Note 2 for the correlation of forms in +ize to forms in +ism.

Paradigm 14 denominal +ism, uncountable (Neo-Latin provenance)

BASE TYPES

- 1 Those nouns, simple or holistic, of Greek or Latin provenance for which the paraphrases work meaningfully:

simple: cube, sex...

holistic: atom, commune, alarm, defeat...

- 2 Comparable bound bases:

femin-, optim-, gigant-, the-, ostrac-...

- 3 Personal surnames, regardless of provenance:

Marx, Mao, Stalin, Darwin, Calvin...

- 4 Group and type names, regardless of provenance:

Whigg, Tory, Quaker, Hellene, Hindu...

- 5 Any of the preceding where (1) a Latin or Greek nominative inflection, or (2) the vowels written as -a or -i/y are deleted:

5.1 Narcissus (narciss-), Thomas (Thom-)...

5.2 Buddha (Buddh-), Nazi (Naz-), academy (academ-)...

DERIVED TYPES

Nouns derivable from words with the following suffixes, whether functional or parts of holisms:

- 1 +al natural, central...
- 2 +an human...
- 3 +ar regular...
- 4 +ary military...
- 5 +ian Prussian, Indian...
- 6 +ic critic, eclectic...
- 7 +ical radical, syndical...
- 8 +id invalid...
- 9 +il(e) infantile...
- 10 +in(e) nicotine...
- 11 +ion union...
- 12 +ional functional, regional...
- 13 +it(e) anti-semitic...
- 14 +ive progressive, negative...
- 15 +mental fundamental...

PARADIGM

<u>n/a</u>	<u>n</u> + <u>ism</u> (stative) (systemic)	<u>n/a</u> → <u>ist</u> (type: (human) (relational)	<u>a</u> + <u>ic</u> (relational) (similative) (depreciative)
------------	--	--	--

EXAMPLES

commun-
communism
communist
communistic

Calvin
Calvinism
Calvinist
Calvinistic

DEFINITIVE PARAPHRASES

- 1 +ism, a stative systemic noun, expressing a kind of state, system, theory, belief, movement, tendency or phenomenon as understood in academic, scientific, political, religious or similar terms:

Atomism is the scientific theory in which atoms are regarded as the smallest particles of physical existence.

Darwinism is the scientific theory which was propounded by Darwin and which proposed the evolution of all life from a common source.

Feminism is the social (and political) movement in which women (= femin-) seek at least equal rights with men.

Catholicism is the religious system which is practised by Catholics and which is characterized by belief in Christ as the Son of God, the Holy Trinity and certain dogmas.

- 2 →ist, both a noun and an adjective:

- 2.1 a human type noun:

He accepts Darwinism.
He is a follower of Darwin. } = He is a Darwinist.
She indulges in escapism. = She is an escapist.

- 2.2 a relational adjective:

These ideas relate to Darwinism. = These ideas are Darwinist.
They are Darwinist ideas.
This principle pertains to feminism. = The principle is feminist.
It is a feminist principle.

- 3 +istic, a relational and simulative adjective used when a clear distinction is needed between the human type and the adjective, and sometimes suggesting the disapproval of the speaker:

These ideas relate to $\begin{cases} \text{humanism.} \\ \text{humanists.} \end{cases} = \begin{cases} \text{These ideas are humanistic.} \\ \text{They are humanistic ideas.} \end{cases}$

His schemes are similar to communism. = His schemes are communistic.
They are communistic schemes.

NOTES

- 1 When +ism is added to a polysyllabic base, the accent generally falls two syllables before the suffix:

humanism, dognatism, communism...

A listable set does not follow this rule and also does not form an adjective in +istic. Examples:

defeat = defeatism = defeatist = *defeatistic

escape = escapism = escapist = *escapistic

alarm = alarmism = alarmist = *alarmistic

- 2 Although there is a close correlation between the base types of +ize and +ism, there is not a one-to-one correlation:

2.1 The existence of itemize does not imply *itemism.

2.2 The existence of Buddhism does not imply *Buddhize.

- 3 Although there is a close correlation between +ism and +ist, there is not a one-to-one correlation:

3.1 The existence of cyclist and violinist (See Paradigm 1, Notes 2.9 and 2.10) does not imply cyclism and violinism.

3.2 The existence of organism and organist does not imply that they are members of the same paradigm.

- 4 With group names (religious and geographical) +ist is not normal invoked:

Hindu = Hinduism = *Hinduist

Quaker = Quakerism = *Quakerist

Mormon = Mormonism = *Mormonist

- 5 There are medico-pathological terms in +ism for which the paradigm does not operate:

adenoidism, gigantism, teratism...


- 6 The suffix +ism can be attached casually to rank-shifted phrases and sentences, especially if they are idiomatic:

6.1 phrase: me-too-ism

6.2 sentence: never-say-die-ism

The paradigm is not usually exploited by such formations.

- 7 The noun and adjective in -ist can contrast with +ian and +ite when attached to personal surnames. In such sets, -ist is neutral, while the others are appreciative and depreciative:

<u>Darwin</u>		<u>Darwinian</u> (appreciative)
		<u>Darwinist</u> (neutral)
		<u>Darwinite</u> (depreciative)

Paradigm 15 denominal +ism, countable (Neo-Latin provenance)

BASE TYPES

- 1 Those bound nominal bases, usually holistic, and disyllabic, of Greek or Latin provenance, for which the paraphrases work meaningfully:

euphem-, anachron-, terat-, atav-, selec-...

- 2 Personal surnames, usually disyllabic, regardless of provenance:

Spooner, Johnson, Kipling, Malaprop...

- 3 Such other personal surnames which, because they are monosyllabic, add +ian to gain bulk:

Shaw → Shawian-, Holmes → Holmesian-...

- 4 Place-related adjectives in:

4.1 -an- Cuban...

4.2 -ian- Canadian, Indian...

4.3 -ish- British, Irish...

4.4 -ic- (1) from +ish British → Britic...

(2) +ic proper Gallio...

PARADIGM

<u>n</u>	<u>n</u> + <u>ism</u> (type: abstract)	<u>a</u> → <u>istic</u> (relational) (attributive)
----------	--	--

EXAMPLES

euphem-
euphemism
euphemistic

anachron-
anachronism
anachronistic

America
American
Americanism
Americanistic

DEFINITIVE PARAPHRASES

- 1 +ism, an abstract type noun, expressing a kind of device, peculiarity, eccentricity, fault, aberration or similar phenomenon relating to literary, stylistic, linguistic, grammatical, dramatic, cultural, pathological or similar subjects:

A euphemism is a linguistic device in which one 'speaks well' (= euphem-) or obliquely of serious or taboo events, such as death.

An anachronism is a literary device in which something occurs in a context where it is 'out of its proper time' (= anachron-).

A teratism is a biological aberration in which a monster (terat-) is formed, or is the monster itself.

An Irishism is an Irish linguistic usage or cultural attitude.

2 →istic, a relational or attributive adjective:

The clock in the play 'Julius Caesar' is anachronistic.

His remarks about his father's death were euphemistic.

NOTES

1 Sometimes a type human noun in →ist occurs:

A euphemist is one who indulges in euphemisms.

2 →istic is not usually invoked for forms based on surnames and place-names:

Irish: Irishism: *Irishistic

Kipling: Kiplingism: *Kiplingistic

3 Although countability marks this paradigm as distinct from Paradigm 14, there is a connection between the two:

Euphemisms are instances of euphemism.

Indianisms may be examples of Indianism.

4 A parallel use of +ese occurs with surnames, place-names and institutions to give type nouns suggesting especially the language, style and ideas of a group:

Journalese is the style of journalists.

Officialese is the language of officials.

This usage is often depreciative, implying that the speaker does not approve.

BIBLIOGRAPHY

Excluding the titles of wordbooks mentioned in the text unless (1) the preface has been quoted from, or (2) extensive use has been made of a particular book

- Abercrombie, David (1965) Studies in Phonetics & Linguistics, Oxford University Press
- _____ (1967) Elements of General Phonetics, Edinburgh University Press
- Abraham, S (1967) 'On a Recent Attempt at a Formal Definition of the Word', Linguistics 35, Sep 1967, 5-6
- Alexander, L G (1967) New Concept English Series, Longman
- Allbrow, K (1970) 'The English Writing System : Notes Towards a Description', Dept of General Linguistics, University College, London
- Barnhart, Clarence L (1968) 'Francis Andrew March : An Appreciation' in March, Thesaurus and Dictionary
- _____ (1973) 'Plan for a Central Archive for Lexicography in English', in McDavid & Duckert
- Bazell, C E (1949) 'On the Problem of the Morpheme', Archivum Linguisticum 1, 1-
- _____ (1952) 'The Correspondence Fallacy in Structural Linguistics', Studies by Members of the English Dept. Istanbul University 3
- _____ (1953) 'Linguistic Form', Istanbul University
- _____ (1954) 'The Sememe', Litera, Istanbul, 17-31
- _____ (1962) 'Meaning and the Morpheme', Word 18, 132-142
- Bloomfield, Leonard (1933a) Language, New York : Holt, Rinehart and Winston
- _____ (1933b) 'The Structure of Learned Words', in Bolton and Crystal, vol 2, 1969
- Bolinger, Dwight L (1948) 'On Defining the Morpheme', Word 4, 18-23
- _____ (1963) 'The Uniqueness of the Word', Lingua 12, 113-136
- _____ (1965) 'The Atomization of Meaning', Language 41, 555-73
- _____ (1968) Aspects of Language, Harcourt Brace and World

- Bolton, W F (1966) ed. The English Language : Essays by English & American Men of Letters, 1490-1839, Cambridge University Press
- Bolton, W F and Crystal, D (1969) eds. vol 2 The English Language : Essays by Linguists & Men of Letters 1858-1964, Cambridge University Press
- Bongers, Herman (1947) The History and Principles of Vocabulary Control Wocopi-Woerden (Holland)
- Boulton, J T (1971) ed. Johnson : The Critical Heritage, Routledge & Kegan Paul, London
- Bright, J A and McGregor, G P (1970) Teaching English as a Second Language, Longman
- Burling, Robbins (1969) 'Cognition and Componential Analysis : God's Truth or Hocus-Pocus?', in Tyler
- Chapman, Robert L (1973) 'On Collecting for the Central Archive', in McDavid & Duckert
- Chesterfield, Lord (1754) two articles in the World Nov/Dec 1754, re-printed in Boulton (1971)
- Chomsky, Noam (1957) Syntactic Structures, Mouton
- _____ (1965) Aspects of the Theory of Syntax, Cambridge, Massachusetts, MIT Press
- Clifford, J L (1955) Young Sam Johnson, McGraw-Hill
- Conklin, Harold C (1969) 'Lexico-graphic Treatment of Folk Taxonomies', in Tyler
- Cook, W A and O'Shea, M V (1914) The Child and his Spelling, Indianapolis
- Cousins, Norman (1968) 'Foreword', in March, Thesaurus and Dictionary
- Crystal, David (1971) Linguistics, Pelican
- Dean, Leonard F, Gibson, Walker and Wilson, Kenneth G (1971) eds. The Play of Language, Oxford University Press
- Dewey, Godfrey (1923) The Relative Frequency of English Speech Sounds, Harvard Studies in Education, Vol. IV, Cambridge, 1923
- Dinneen, Francis P (1967) An Introduction to General Linguistics, Holt Rinehart and Winston Inc
- Dutch, Robert A (1962) ed. Roget's Thesaurus, Longman

- Eaton, Helen S (1934) Comparative Word Frequency List, International Auxiliary Language Association, New York
- Egan, Rose F (1942) 'Survey of the History of English Synonymy', in the Merriam-Webster Dictionary of Synonyms, my edition, 1968
- Eldridge, R C (1911) Six Thousand Common English Words, Niagara Falls
- Emblen, D L (1970) Peter Mark Roget, The Word and the Man, Longman
- Entwhistle, William J (1953) Aspects of Language, Faber, London
- Faucett, Lawrence and Maki, Itsu (1932) Study of English Word-Values Statistically Determined from the Latest Extensive Word-Counts, Tokyo
- Finkenstaedt, Thomas and Wolff, Dieter (1970) A Chronological English Dictionary, Heidelberg, Carl Winter, Universitätsverlag
- Finkenstaedt, Thomas, Leisi, Ernest and Wolff, Dieter (1973) Ordered Profusion, Heidelberg, Carl Winter, Universitätsverlag
- Firth, J R (1951) Papers in Linguistics, 1934-1951, Oxford
- Frake, Charles O (1969) 'The Ethnographic Study of Cognitive Systems', in Tyler
- Fries, Charles C and Traver, A Aileen (1950) English Word Lists A Study of their Adaptability for Introduction : the George Wahr Publishing Co, Ann Arbor, Michigan
- Gimson, A C (1973) 'Phonology and the Lexicographer', in McDavid and Duckert
- Gleitman, Lila R and Henry (1970) Phrase and Paraphrase, W W Norton, New York
- Greene, D J (1970) Samuel Johnson, Twayne Publishers Inc., New York
- Grove, Victor (1949-50) The Language Bar, Routledge & Kegan Paul
- Haas, William (1973) review of Lyons, Introduction to Theoretical Linguistics, in Journal of Linguistics 9:1, Feb
- Halliday, M A K (1961) 'Categories of the Theory of Grammar', in Word 17:3, Dec
- _____ (1970) A Course in Spoken English : Intonation, Oxford University Press
- Harris, Zellig S (1942) 'Morpheme Alternants in Linguistic Analysis', Language 18, 169-181
- _____ (1945) 'Discontinuous Morphemes', Language 21, 121-128
- _____ (1946) 'From Morpheme to Utterance', Language 22, 161-184

- _____ (1951) 'Methods in Structural Linguistics, The University of Chicago Press
- Hill, Archibald A (1958) Introduction to Linguistic Structure, Harcourt, Brace and World Inc
- _____ (1969) ed. Linguistics Today, Basic Books, New York
- Hockett, Charles F (1947a) Review of Nida's Morphology, Language 23, 274-285
- _____ (1947b) 'Problems of Morphemic Analysis' Language 23, 321-344
- _____ (1947c) 'Peiping Morphophonemics' Language 26, 63-85
- _____ (1954) 'Two Models of Grammatical Description', Word 10, 210-231
- Hogben, Lancelot (1964) The Mother Tongue, Secker and Warburg
- Hornby, A S, Gatenby, E V and Wakefield, H (1948) The Advanced Learner's Dictionary of Current English, Oxford University Press
- Householder, Fred W and Saporta, Sol (1962) eds. Problems in Lexicography, Bloomington (Indiana), and Mouton, 2nd edn, with additions & corrections, 1967
- Hulbert, James R (1955) Dictionaries British & American, André Deutsch
- Hymes, Dell H (1969) discussion of Burling, in Tyler
- _____ (1971) ed. Pidginization & Creolization in Languages, Cambridge University Press
- Jespersen, Otto (1954) A Modern English Grammar on Historical Principles, George Allen & Unwin (Preface dated 1942), Vol VI, ch 1
- Johnson, Samuel (1747) Plan for a Dictionary of the English Language, London
- _____ (1755) Preface to A Dictionary of the English Language, London, reprinted in Bolton (1966)
- Jones, Daniel (1917) An English Pronouncing Dictionary, London
- Katz, J J and Fodor, J A (1963) 'The Structure of a Semantic Theory', Language 39, 170-210
- _____ and Postal, P M (1964) An Integrated Theory of Linguistic Descriptions (Research Monographs 26), Cambridge Massachusetts, MIT Press
- _____ (1966) The Philosophy of Language, New York, Harper and Row

- Knowles, J (1904) The London Print System of Reading for the Blind, London
- Kuhn, Thomas (1962) The Structure of Scientific Revolutions, University of Chicago Press
- Lakoff, George (1973) 'Lexicography & Generative Grammar I : Hedges and Meaning Criteria', in McDavid & Duckert
- Lamb, Sidney M (1969) 'Lexicology and Semantics', in Hill
- Lehmann, W P (1973) 'On the Design of a Central Archive for Lexicography in English', in McDavid & Duckert
- Lejnieks, Valdis (1967) 'The System of English Suffixes', Linguistics 29, Feb, 80-103
- Lloyd-James (1935) The Broadcast Word, London
- Lyons, John (1968) Introduction to Theoretical Linguistics, Cambridge University Press
- _____ (1970) New Horizons in Linguistics, Pelican
- Malkiel, Yakov (1968) Essays on Linguistic Themes, Basil Blackwell, Oxford
- _____ (1973) discussion statement, in McDavid and Duckert
- Marb , Alan and McArthur, Tom (1975) Natural Speech : A course in the Rhythm and Intonation of Natural English, Department of Educational Studies, University of Edinburgh
- March, Francis A (1902) ~~1968~~ Thesaurus and Dictionary of the English Language, Doubleday, New York (my edition 1968)
- Marchand, Hans (1960) ~~and 1969~~ The Categories and Types of Present-Day English Word-Formation, Verlag C H Beck, Munchen (second edition, 1969)
- Mathews, M M (1933) A Survey of English Dictionaries, New York, Russell & Russell (1966 re-issue)
- Matthews, P H (1965a) 'The inflectional component of a word-and-paradigm grammar' Journal of Linguistics 1, 139-71
- _____ (1965b) 'Some concepts in word-and-paradigm morphology', Foundations of Language 1, 268-89
- _____ (1967a) 'Latin', Lingua 17, 163-81
- _____ (1967b) 'Review of Chomsky' (1965), Journal of Linguistics 3

- _____ (1970) 'Recent Developments in Morphology', in Lyons
McArthur, Tom (1972a) Building English Words, 'Patterns of English', Collins
_____ (1972b) Using English Prefixes and Suffixes, 'Patterns of English'
Collins
_____ (1972c) Using Compound Words, 'Patterns of English', Collins
_____ (1973) Using Phrasal Verbs, 'Patterns of English', Collins
_____ and Atkins, Beryl (1974) Dictionary of Phrasal Verbs and Their
Idioms, Collins
McDavid, Raven I, and Duckert, Audrey R (1973) Lexicography in English, New York
Academy of Sciences
Miller, R L (1966) 'The Word and its Meaning', Linguistics 28 Dec, 90-96
Murray, James A H (1888) 'General Explanations' to Vol I of the New English
Dictionary; also in Bolton & Crystal
_____ (1900) The Evolution of English Lexicography : The Romanes
Lecture, 22 June 1900, Oxford
Newman, Stanley S (1935) 'English Suffixation : a Descriptive Approach', Word 4,
1948, 24-36
Nida, Eugene A (1948a) 'The Analysis of Grammatical Constituents' Language 24,
168-178
_____ (1948b) 'The Identification of Morphemes', Language 24, 414-441
_____ (1949) Morphology : the Descriptive Analysis of Words (2nd edn)
reprint 1956, Ann Arbor, the University of Michigan Press
Ogden, C K (1930) The Basic Dictionary, Psyche Miniatures, London
_____ (1933) The Basic Words. a Detailed Account of Their Uses, Psyche
Miniatures, London
_____ (1935) Counter-Offensive : An Exposure of Certain Misrepresentations
of Basic English, the Orthological Institute, Cambridge
_____ (1940) The General Basic English Dictionary, Evans
_____ and Richards, I A (1923) The Meaning of Meaning, 1946 edition,
Routledge & Kegan Paul
Onions, C T (1966) The Oxford Dictionary of English Etymology, Oxford University
Press

- Palmer, Harold E (1930) Interim Report on Vocabulary Selection, Institute for Research in English Teaching, Tokyo
- _____ (1931) Second Interim Report on Vocabulary Selection, Institute for Research in English Teaching, Tokyo
- _____ West, Michael and Faucett, Lawrence (1936) Interim Report on Vocabulary Selection, King & Son, London (Carnegie Conference Report)
- _____ and Hornby, A S (1937) Thousand-Word English : What it is and what can be done with it, Harrap
- Quirk, Randolph (1962 : 1968 2nd edn) The Use of English, Longman, including Warburg, Jeremy 'Notions of Correctness'
- _____ and Greenbaum, S (1973) A University Grammar of English, Longman
- Read, Allen W (1973) 'The Social Impact of Dictionaries in the United States', in McDavid and Duckert
- Robertson, Stuart and Cassidy, Frederic G (1954) 'Changing Meanings and Values of Words', in Dean, Leonard F et al (1971)
- Robins, R H (1959) 'In Defence of W P', in Transactions of the Philological Society, 116-44
- Roget, Peter Mark (1853:1962) Thesaurus of English Words and Phrases, ed. Robert A Dutch
- Ross, Alan S C (1958) Etymology, with Special Reference to English, André Deutsch
- Sapir, Edward (1921) Language, An Introduction to the Study of Speech New York : Harcourt Brace & World Inc
- Saussure, Ferdinand de (1916) Cours de Linguistique Generale, 5th edn, Paris : Payot, 1955
- Scott, F S and Bowley, Brockett, Brown, Goddard (1968/1970) English Grammar, A Linguistic Study of its Classes and Structures : Heinemann, London
- Skeat, Walter W (1882) A Concise Etymological Dictionary of the English Language Oxford University Press (1965 impression)
- Sledd, James and Ebbitt, Wilma R (1962) Dictionaries and THAT Dictionary : A

Casebook of the Aims of Lexicographers and the Targets of Reviewers

Scott Foresman & Co

- _____ (1972) 'Dollars and Dictionaries', in Weinbrot
- Stankiewicz, Edward (1962) 'The Interdependence of Paradigmatic and Derivational Patterns', Word 18, 1-22
- Starnes, D T and Noyes G E (1946) The English Dictionary from Cawdrey to Johnson, 1604-1755, Chapel Hill, the University of North Carolina Press
- Strang, Barbara M H (1962) Modern English Structure, Arnold
- _____ (1970) A History of English, Methuen
- TCLP 4 Transactions of the Linguistic Circle of Prague No 4, 1931
- Thorndike, Edward L (1921a) The Teacher's Word Book, Teachers College, Columbia University
- _____ (1921b) 'Word Knowledge in the Elementary School', Teachers College Record XXII
- _____ (1932) A Teacher's Wordbook of the 20,000 Words Found Most Frequently and Widely in General Reading for Children and Young People, Teachers College, Columbia University
- _____ and Lorge, Irving (1944) The Teacher's Word Book of 30,000 Words, Bureau of Publications, Teachers College, Columbia University, New York
- _____ (1948) Thorndike English Dictionary, The English University Press (Hodder & Stoughton)
- Trager, George L (1951) Review of Nida's Morphology in International Journal of American Linguistics 17 : 126-131
- Twadell, W F (1973) 'Lexicography and People', in McDavid & Duckert
- Tyler, Stephen A (1969) ed. Cognitive Anthropology, Holt Rinehart & Winston
- Vendryes, Joseph (1923) Le Langage, Paris
- Warburg, Jeremy (1968) 'Notions of Correctness', in Quirk
- Webster, Noah (1807) 'Letter to Dr David Ramsay', in Boulton, 125-140
- _____ (1828) preface to An American Dictionary of the English Language reprinted in part in Sledd and Ebbitt

Weekley, Ernest (1924) 'On Dictionaries', in Sledd & Ebbitt

Weinbrot, Howard D (1972) ed. New Aspects of Lexicography, Southern Illinois University Press and Feffer & Simons, London & Amsterdam

Weinreich, Uriel (1962) 'Lexicographic Definition in Descriptive Semantics', in Householder & Spota

_____ (1964) review of Webster's Third, in International Journal of American Linguistics 30

_____ (1971) in Steinberg, Danny and Jakobovits, Leon A, eds. Semantics: an interdisciplinary reader in philosophy, linguistics and psychology, Cambridge University Press

West, Michael (1927 etc) The New Method Readers, Longman

_____, Swanson, E and others (1934) A Critical Examination of Basic English, University of Toronto Press

_____ (1935) New Method Dictionary, Longman

_____ (1953) A General Service List of English Words, Longman

_____ (1965) An International Reader's Dictionary, Longman

Whitehall, Harold (1958) 'The Development of the English Dictionary', in Dean, Leonard F et al (1971)

Widdowson, H G (1968) 'The Teaching of English Through Science' in Dakin, Julian Tiffen, Brian and Widdowson H G Language in Education, Oxford University Press

Wilks, Y A (1972) Grammar, Meaning and the Machine Analysis of Language, Routledge & Kegan Paul

Wright, Thomas (1857) A Volume of Vocabularies, ed. by R P ^uWilcker as Anglo-Saxon and Old English Vocabularies (1884), London

^uWilcker see Wright

Zgusta, Ladislav (1971) Manual of Lexicography, Mouton

_____ (1973) 'Lexicology : Generating Words', in McDavid & Duckert

Zirmunskij, V M (1966) 'The Word and its Boundaries', Linguistics 27, Nov, 65-9